

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE ECONOMIA
MONOGRAFIA DE BACHARELADO

**A qualidade das informações de Cor ou Raça na RAIS:
um estudo comparativo com o CENSO Demográfico de
2010**

GUILHERME CARDOSO PORTELA CÂMARA
MATRÍCULA nº109023789

ORIENTADOR: PROF. MARCELO JORGE DE PAULA PAIXÃO.

Fevereiro 2015

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE ECONOMIA
MONOGRAFIA DE BACHARELADO

**A qualidade das informações de Cor ou Raça na RAIS:
um estudo comparativo com o CENSO Demográfico de
2010**

GUILHERME CARDOSO PORTELA CÂMARA
MATRÍCULA nº 109023789

ORIENTADOR: PROF. MARCELO JORGE DE PAULA PAIXÃO.

Fevereiro 2015

As opiniões expressas neste trabalho são de exclusiva responsabilidade do autor.

AGRADECIMENTOS

Em primeiro lugar, agradeço o apoio do meu irmão Leonardo que empreendeu minuciosa correção do texto. Agradeço também ao seu incondicional amparo e camaradagem. A toda a família – àqueles próximos ou distantes – um caloroso abraço.

Ao LAESER sem o qual não teria sido introduzido nessa importante agenda de pesquisa, nem esse trabalho possível. Ao meu orientador pela pronta recepção e correção dessa monografia.

RESUMO

No presente trabalho comparamos os dados da Relação Anual de Informações Sociais (RAIS), do Ministério do Trabalho e Emprego (MTE) e do Censo Demográfico de 2010, do Instituto Brasileiro de Geografia e Estatística (IBGE) de modo a identificar a presença de diferenças significativas na maneira como estas fontes de dados capturam a informação de cor ou raça dos trabalhadores, bem como a desigualdade salarial associada a esta variável. Executa-se uma revisão do tema, identificando como traços característicos das diferenças metodológicas – isto é, de heteroatribuição e autoatribuição de pertença racial – a tendência do primeiro de superestimar a população branca e a subestimar magnitude da discriminação racial.

A análise dos dados verifica a presença de superestimação da população branca em algumas das regiões geográficas, mas não verifica maior discriminação racial na RAIS em relação ao Censo demográfico. Os modelos econométricos utilizados comprovam diferenças estatisticamente significativas na captura da desigualdade racial entre as duas pesquisas, além de diferenças em termos de produtividade associadas às variáveis educacionais.

ÍNDICE

INTRODUÇÃO	8
CAPÍTULO I – REGISTROS ADMINISTRATIVOS E PESQUISAS AMOSTRAIS.....	10
CAPÍTULO II - CARACTERIZAÇÃO DA RAIS E DIFERENÇAS COM O CENSO	15
II.1 – DA RAIS E SUA ÁREA DE COBERTURA	15
II.2 – PERÍODO DE REFERÊNCIA	17
II.3 – RENDA DO TRABALHO.....	18
II.4 – VARIÁVEL DE COR OU RAÇA	23
II.5 – OUTRAS CONSIDERAÇÕES.....	29
CAPÍTULO III – DISCUSSÃO DOS DADOS QUANTITATIVOS	30
III.1 – DADOS DAS UNIDADES FEDERATIVAS	30
III.2 – DADOS POR NÍVEL EDUCACIONAL	36
CAPÍTULO IV – APLICAÇÃO DE MODELOS DE EQUAÇÕES DE SALÁRIOS	40
IV.1 – EQUAÇÕES DE SALÁRIOS NO CENSO E RAIS.....	40
IV.2 – MODELO DE EQUAÇÕES DE SALÁRIOS EM UMA AMOSTRA MISTA	43
CAPÍTULO V – DECOMPOSIÇÃO DE OAXACA PARA AS EQUAÇÕES DE SALÁRIOS	46
V.1 – DECOMPOSIÇÃO DE OAXACA.....	46
V.2 – RESULTADOS DA DECOMPOSIÇÃO	47
CONCLUSÃO	51
REFERÊNCIAS BIBLIOGRÁFICAS	53

ÍNDICE DE GRÁFICOS

Gráfico 1 - Curva de Lorenz para a Renda do Trabalho/Vínculo do CENSO Demográfico e da Relação Anual de Informações Sociais, Brasil, 2010	22
---	----

ÍNDICE DE TABELAS

Tabela 1 - Composição por UFs e distribuição por Sexo e Cor ou Raça da população ocupada com carteira assinada por UFs. Comparação entre Censo e RAIS, 2010	31
---	----

Tabela 2 - Renda Nominal Média do trabalho por UFs e Cor ou Raça da população ocupada com carteira assinada por UFs. Comparação entre Censo e RAIS, Brasil, 2010 ...	34
--	----

Tabela 3 - Composição de maior nível educacional alcançado por Sexo e Cor ou Raça da população ocupada com carteira assinada. Comparação entre Censo e RAIS, Brasil, 2010	37
---	----

Tabela 4 - Renda Nominal Média do trabalho do trabalho por maior nível educacional alcançado, Sexo e Cor ou Raça da população ocupada com carteira assinada. Comparação entre Censo e RAIS, Brasil, 2010	38
--	----

Tabela 5 - Resultados de um modelo de equações de salários minceriana por MQO. Comparação entre Censo e RAIS, Brasil, 2010.....	42
---	----

Tabela 6 - Modelo de equações de salários por MQO em uma amostra mista. Comparação entre Censo e RAIS, Brasil	44
---	----

Tabela 7 - Resultados de um modelo de equações de rendimentos por MQO para quatro populações. Censo 2010, Brasil.....	48
---	----

Tabela 8 - Resultados de um modelo de equações de rendimentos por MQO para quatro populações. RAIS, 2010, Brasil	49
--	----

Tabela 9 - Resultados da decomposição de Oaxaca para três grupos em relação à população masculina branca, em log natural e (%). Comparação entre Censo e RAIS, Brasil, 2010	50
---	----

INTRODUÇÃO

Dados são escassos e por isso valiosos. Embora às vezes mais pareça que nos afogamos neles, o aparecimento de novas fontes de dados representa acontecimento excitante, abrindo perspectivas de novas descobertas e desenvolvimentos da ciência (sendo esses ainda mais escassos). Uma oportunidade desse tipo se apresenta com a recente liberação, sem restrição ao público, de dados da Relação Anual de Informações Sociais – a RAIS – e, mais recentemente ainda, da variável de cor ou raça do trabalhador. É claro também que fontes de dados não comprovadas não podem ser tomadas pelo seu valor de face. Histórico de uso comprovado é igualmente valioso, e há certo aspecto moral no uso de certas bases de dados: muito depende da autoridade e integridade das instituições que as produzem. Não se pode esperar dos usuários conhecimento pleno da metodologia e das condições de coleta dos dados (embora seja recomendável), meio pelo qual poderiam tirar suas próprias conclusões sobre a adequação delas.

Destas considerações parte o propósito do presente trabalho: validar ou não a adequação da RAIS para estudos utilizando a variável de cor ou raça. Para isso, empregaremos o Censo Demográfico de 2010, já que se admite a qualidade da pesquisa para finalidades de entendimento da dinâmica do mercado de trabalho brasileiro (IBGE, 2013).

A desigualdade racial no Brasil é componente dessa dinâmica: grave fato, uma vez que boa parte dos trabalhadores é discriminada unicamente pelas suas características superficiais e não por mérito ou zelo no desempenho de seu emprego (HASENBALG, 1979; VALLE SILVA, 1980; OLIVEIRA, PORCARO & SILVA 1983; SOARES, 2000; MARTINS, 2003a, 2003b; TELLES, 2003). Dado que a redução da desigualdade de renda brasileira é objetivo de políticas públicas, para alcançá-lo deve-se, necessariamente, empregar políticas acessórias e consistentes, inclusive as de combate à discriminação racial nas suas múltiplas formas. Daí a importância da pesquisa acadêmica e de dados concretos sobre o tema.

A resposta do problema proposto se dá em cinco capítulos mais a conclusão. No primeiro deles, discutiremos as diferenças entre pesquisas amostrais e registros administrativos, o dilema do *trade off* no uso de fontes alternativas, histórico literário deste tipo de comparação e a utilização do Censo como parâmetro de comparação.

No segundo capítulo trataremos, em cinco seções, das diferenças metodológicas entre as duas pesquisas, a saber: da área de cobertura, do período de referência, da captura dos rendimentos do trabalho, da variável de cor ou raça e outras compatibilizações. Ao fim obteremos as especificações de um corte transversal das duas pesquisas que permitirá uma comparação mais aproximada, além de uma visão dos eventos que podem causar incongruências no processo comparativo.

No terceiro capítulo, nos deteremos em uma discussão sobre os dados quantitativos comparados, tendo em mente a revisão feita no capítulo anterior. Veremos os dados de composição por unidades da federação e por níveis educacionais. Chegaremos, a partir deste processo, a importantes resultados de como os dados da RAIS se desviam em relação ao nosso parâmetro normativo.

No capítulo seguinte desenvolveremos dois modelos de equações mincerianas clássicas a fim de isolar os efeitos das características raciais do indivíduo sobre a renda. A partir disso poderemos identificar a presença de discriminação racial nas duas bases de dados e sua magnitude. Na primeira seção serão descritos os modelos individualmente produzidos para as bases e seus resultados. Na segunda, o modelo anterior será transformado com o objetivo de estimar as diferenças entre as duas fontes em uma amostra mista. Este modelo vai permitir um teste conclusivo sobre essas estimativas.

Por fim, no capítulo quinto será executado a decomposição dos modelos referidos acima com base em Oaxaca (1974) e determinaremos a importância da discriminação racial e de gênero na composição da renda do trabalho. À guisa de conclusão, serão propostos algumas recomendações e desafios.

CAPÍTULO I – REGISTROS ADMINISTRATIVOS E PESQUISAS AMOSTRAIS

O trabalho com bases de dados estatísticos e indicadores sociais raramente é trivial. A sugestão de facilidade, oferecida pelos modernos meios computacionais e *softwares* estatísticos dos quais dispomos, conquanto muito bem-vindos, não elimina as decisões cruciais que dependem unicamente do pesquisador. Pois é fato: cabe ao arbítrio dele a escolha entre fontes de informação diferentes para os mesmos objetos estudados. A essas complicações, somam-se a escolha entre marcos teóricos diferentes e a complexa relação entre a teoria e a própria observação dos fatos – pois estes também são produzidos. Para cada pesquisa, registro administrativo ou histórico, há um sem-número de elementos a serem considerados. Por exemplo: a forma verbal das perguntas, o desenho amostral, as ideologias inculcadas nos entrevistadores e entrevistados. Dessa forma, considerando os custos envolvidos na coleta e produção de informações estatísticas, sempre há deficiências e omissões. Para os mesmos objetos estudados, comparando-se fontes de dados estatísticos intercambiáveis, os resultados frequentemente são incongruentes, mas mesmo assim fidedignos. Há considerável *trade off* nas escolhas disponíveis. Quê fazer? Não há teoria que nos informe – a priori – qual a melhor decisão a ser tomada na seleção de indicadores sociais e suas fontes alternativas (JANNUZZI, 2001).

Postas essas considerações, o caso da variável cor ou raça é bastante peculiar. Em virtude da relativa escassez e novidade dessa variável em diversas pesquisas e registros administrativos¹, a escolha entre fontes de dados estatísticos é, em muitos casos, trivial. Sabemos que, em quase sua totalidade, as bases que amparam estudos acadêmicos do tema são as pesquisas domiciliares oficiais do Instituto Brasileiro de Geografia e Estatística (IBGE), ou seja: os Censos Demográficos decenais, a Pesquisa por Amostra de Domicílios (PNAD) e a Pesquisa Mensal de Emprego (PME). Isso não é coincidência. Tais pesquisas são as que inquiram há mais tempo acerca dessas características², uma vantagem por si só no trabalho acadêmico. Essa exclusividade se manifesta claramente no histórico das pesquisas de discriminação racial no mercado de trabalho brasileiro. Destacamos os trabalhos pioneiros no tema: Hasenbalg (1979), Valle Silva

¹ Para conferir um sumário de fontes de dados estatísticos raciais, consulte Osório (2003) e Paixão e Rossetto (2011).

² Nos Censos Demográficos, com várias interrupções desde 1872, permanecendo inalterado desde 1991. Na PNAD, aparece em categoria especial em 1976, nas demais edições desde 1987. Na PME, em sua série atual, esteve sempre presente.

(1980), Oliveira, Porcaro e Silva (1983), suas continuações e outras mais recentes como Soares (2000) e Martins (2003a, 2003b), assim como, é claro, os trabalhos do LAE-SER³.

Essa preferência, originada no desenvolvimento histórico do sistema de informações estatísticas oficiais é, contudo, plenamente justificada. Embora confira certa limitação sobre o escopo do trabalho acadêmico no tema da discriminação racial, as dificuldades associadas acabam por encaminhar o pesquisador para essas pesquisas domiciliares oficiais. Afinal, o método de identificação étnica-racial internacionalmente recomendado é o por autoatribuição. Em particular no Brasil, em que há maior ambiguidade na definição prática do que constitui raça, a única subjetividade segura de respostas talvez seja a dos entrevistados. A consistência de resposta entre registros administrativos e pesquisas amostrais – adotando métodos atributivos diferentes – não é a mesma que, por exemplo, nos Estados Unidos, onde as normas raciais são mais reificadas (NRC, 2004)⁴.

É bom esclarecer que, grosso modo, há dois tipos de fontes de dados estatísticos que subsidiam pesquisas do mercado de trabalho brasileiro: as pesquisas amostrais e os registros administrativos. As primeiras consistem habitualmente nas pesquisas domiciliares do IBGE, coletadas de uma seleção aleatória de indivíduos em um plano amostral; o tratamento estatístico dos dados garante que a expansão da amostra represente um retrato fiel do objeto estudado. Os registros administrativos, por sua vez, têm como um dos seus mais importantes representantes os dados do Ministério do Trabalho e Emprego (MTE)⁵. Neste último caso, os dados são coletados mediante informação compulsória das empresas, gerando um registro cujo propósito é o de monitoramento administrativo de benefícios legais aos seus empregados.

As pesquisas amostrais em domicílios permitem um conjunto rico de informações, com questionário específico executado por equipes treinadas – fato que só é possível em amostras pequenas. A depender do desenho amostral, podem possuir melhor

³ Acrônimo de Laboratório de Análises Econômicas, Históricas, Sociais e Estatísticas das Relações Raciais, vinculado ao Instituto de Economia da Universidade Federal do Rio de Janeiro (IE-UFRJ).

⁴ Sobre o tema, Heteroatribuição e Autoatribuição de pertença, ver Telles e Lim (1998), Valle Silva (1999) e Osório (2003).

⁵ A saber, a RAIS – já mencionada – e o Cadastro Geral de Empregados e Desempregados (CAGED, MTE).

representatividade e confiabilidade que registros de carácter censitário, e a um custo inferior. Por fim, essas pesquisas conseguem captar o grande contingente de trabalhadores não legalizados e, por isso, invisíveis às exigências legais de registro (RAMOS, 2007). São instrumentos eficientes para a apreensão da conjuntura e estrutura econômica brasileira. Seu histórico de uso chancela essa conclusão.

Ao lidarmos com o atributo de cor ou raça em registros administrativos, há uma clara deficiência qualitativa. A dificuldade de se obter respostas coerentes nesses questionários⁶ resulta frequentemente em péssimos resultados de pesquisa. Por exemplo, conta-nos Osório (2003): no SIAPE (2003), somente 40% dos funcionários preencheram o quesito em seus formulários; no SIM (2000), 16% dos óbitos não registraram a variável; no SINASC (2000), 13% dos declarados nascidos não tiveram a característica registrada. Na RAIS de 2010, 19,2% dos vínculos empregatícios não registraram cor ou raça (PAIXÃO E ROSSETTO, 2011). Entretanto, é importante esclarecer que, apesar das deficiências, esses sistemas de informações estão sujeitos a considerável progresso na captura das referidas. A introdução de meios eletrônicos de registro e orientação disciplinar dos informantes produzem efeitos sensíveis na qualidade dos dados obtidos⁷.

A despeito dessas considerações, o aparecimento recente de novas fontes de dados estatísticos para o atributo de Cor ou Raça representa novas oportunidades e pautas de pesquisa sobre o tema. Mesmo se, em visão superficial, as incongruências nos resultados estatísticos se revelem consideráveis – conforme indica um estudo anterior (PAIXÃO *et al.*, 2012) – investigações detalhadas dessas fontes estatísticas podem ser de grande proveito. Como foi dito, a utilização de documentos diversos depende da consideração do pesquisador acerca das vicissitudes e *trade off* oferecidos pelos instrumentos à mão. Para tanto, esses registros administrativos possuem qualidades únicas que justificam sua adoção. Talvez o melhor argumento para seu emprego, ao menos em caráter exploratório, é o potencial para melhorias na coleta e do tratamento dos dados. O *feed-*

⁶ É difícil inclusive em entrevistas com pessoal treinado, como relatam Piza e Rosenberg (1999).

⁷ Por exemplo, no ano de 2012, aproximadamente 5,6% dos óbitos registrados pelo SIM não tiveram cor ou raça declarada. Melhorias na coleta de informações em registros administrativos podem amiúde representar verdadeiras quebras em uma série histórica.

No caso da RAIS, os problemas de omissões nos formulários a serem declarados têm sido paulatinamente reduzidos através dos meios eletrônicos. “Tais problemas (omissão de formulários) têm sido, a cada ano, consideravelmente reduzidos com a recepção das declarações – desde o ano base de 2000 – apenas em meio magnético e pela Internet; nas quais são feitas as primeiras críticas na preparação e coleta dos dados” (MTE, 2014).

back que o pesquisador pode oferecer ao poder público em direção a tais melhorias é valioso.

Conquanto restrito ao seu lado formal, a RAIS possui enorme potencial para a compreensão das estruturas do mercado de trabalho brasileiro (JANNUZZI, 1994), embora seja preterido por pesquisas amostrais. Em tese, se há esse potencial, haverá também para a compreensão da discriminação racial no mercado de trabalho. Para Ária e Cordeiro (1990), o maior atrativo dessa relação anual se dá pelo caráter censitário do mesmo. Assim como consistência é uma qualidade desejável, o número de observações sempre o é. Habilita pautas de pesquisa para estudos locais ou fortemente desagregados, algo impossível de se atingir em pesquisas como a PNAD sem incorrer em considerável perda de significância estatística. Aquilo que Goldberger (1991) definiu jocosamente como “micronumerosidade” não é empecilho, permitindo que o pesquisador não abandone resultados ou opere tratamento excessivo destes.

Em conjunto com a possibilidade de se trabalhar com uma quantidade extensa de categorias ocupacionais e setoriais por meio dos códigos da CBO e CNAE, julga-se que a RAIS pode ser de grande utilidade para as pesquisas pautadas em teses consensuais de “racismo à brasileira” (TELLES, 2003). Quer dizer, hipóteses sobre a existência de barreiras à ascensão ocupacional de populações pretas e pardas no Brasil, assim como de discriminação crescente em níveis sociais superiores (SOARES, 2000; CORSEUIL E COELHO, 2002), podem ser livremente testadas (se a qualidade dos dados for verificável). Isto é especialmente verdadeiro uma vez que é possível trabalhar com dados longitudinais e não somente com cortes transversais. Também é possível avaliar as diferenças, raramente observadas, de critérios de autoatribuição de pertença e heteroatribuição para atributos raciais.

A literatura sobre o tema da comparação entre a RAIS e as pesquisas amostrais do IBGE é escassa, mas suficiente. Empregando a variável cor ou raça, somente Paixão *et al.* (2012) o fizeram, mas sem acesso robusto dos microdados. De Negri *et al.* (2001) e Corseuil e Santos (2002), compartilhando de métodos semelhantes e em comparação com a PNAD, observaram que haviam algumas semelhanças nos dados e nos preços salariais implícitos para os parâmetros de *dummies* de sexo, idade e região geográfica (obtidos por meio de MQO e *bootstrapping*), ao passo que os parâmetros das variáveis de

dummies educacionais apresentavam diferenças consideráveis. A distribuição de renda era semelhante, observável através dos coeficientes de variação e da composição dos grupos de faixa de renda (assim, de modo pouco sofisticado).

Seguirá nesta monografia uma análise comparativa nos moldes desses trabalhos. Contudo, desta vez serão comparados os dados da RAIS com os do Censo Demográfico de 2010. Apesar do caráter anual da RAIS e sua afinidade natural com a PNAD, aproveitaremos a proximidade deste trabalho com outros do LAESER utilizando os dados do Censo Demográfico. Afinal, queremos produzir um critério permanente pelo qual se possa julgar a validade dos dados da RAIS. O Censo, com seu amplo tamanho de amostra, possui características únicas na distribuição de renda, além de um desenho amostral menos enviesado que o da PNAD (SOUZA, 2013). Assim, na comparação nacional e entre unidades federativas – em especial nas menores – esperamos obter resultados melhores. Evitamos até agora falar das características específicas da RAIS. Essas características, como diferem do Censo de 2010 e como podem ser compatibilizadas, serão discutidas no capítulo seguinte.

CAPÍTULO II – CARACTERIZAÇÃO DA RAIS E DIFERENÇAS COM O CENSO

Neste capítulo serão discutidos aspectos metodológicos das duas bases de dados que constituem o objeto de comparação da presente monografia: a RAIS e o Censo de 2010. Perfilam-se em sequência cinco seções: da RAIS e sua área de cobertura; o período de referência; metodologia da informação de rendimentos; a variável Cor ou Raça; e outras considerações.

II.1 – DA RAIS E SUA ÁREA DE COBERTURA

A RAIS é um registro administrativo anual de âmbito nacional promulgado em 1975, por força do Decreto nº 76.900/75. Constitui uma fonte alternativa de dados sociais e demográficos sobre o mercado de trabalho formal. As informações coletadas são compostas dos formulários enviados ao Ministério do Trabalho e Emprego (MTE) por todas as empresas registradas no Cadastro Nacional de Pessoas Jurídicas (CNPJ), incluindo aí as repartições públicas, estatais e autarquias. Os dados dispõem de informações do estoque e de movimentação de empregados. Em destaque, apresentam informações sobre admissões, desligamentos, escolaridade, carga horária, rendimento, setor de atividade, código de ocupação dos empregados, tipo de vínculo, sexo, nível educacional. Além disso, o tratamento dos dados confere à RAIS a possibilidade de desagregação geográfica até ao nível municipal e sub-regiões de algumas metrópoles.

Januzzi (1995) afirma que diversos trabalhos, como Azevedo (1985), Sabóia e Tolipan (1985) e Ipardes (1987), demonstram que a RAIS constitui um “retrato parcial do mercado de trabalho” (p. 205), isto é, não representa a totalidade do mesmo. Isso se dá em consequência, como já foi dito, do fato de lidar somente com os vínculos de trabalho legais, sendo excluídos ainda a informação de empregadores, sócios, cargos eletivos, estagiários regidos pela Portaria MTPS nº 1.002/1967 e Lei nº 11.788/2008, além de empregados domésticos regidos pela Lei nº 11.324/2006 (MTE, 2011; PAIXÃO et al., 2012). Observemos que todos esses dados são, por outro lado, capturados pelos Censos Demográficos. De fato, essa é uma grande lacuna da RAIS. Porém, para a população estudada – isto é, formada pelos empregados com vínculo legal de trabalho –, a cobertura da RAIS é excelente, superior a 97%. Como veremos, ela goza de uma boa

representatividade regional⁸: não há nenhuma discrepância severa em relação ao Censo Demográfico de 2010. Corseuil e Santos (2002), em comparação com a PNAD de 1998, constata o mesmo. Em parte, as conclusões desses quatro primeiros trabalhos citados são datadas e não dão conta dos imensos avanços na coleta e cobertura dos formulários das empresas brasileiras, em especial a partir do ano de 2000, quando a coleta das informações tornou-se inteiramente por meio eletrônico.

Ressalva-se que a função legal desempenhada pela RAIS é de caráter majoritariamente administrativo, principalmente para efetuar o controle da Lei dos 2/3⁹, do controle relativo ao FGTS e Previdência Social, para viabilização do pagamento do Abono Salarial previsto na Constituição Federal e da composição do CNIS. Sua função, como “um censo administrativo sobre o mercado de trabalho” (JANNUZZI, *op. cit.*, p. 204), é secundário ao seu papel administrativo e, dessa forma, não apresenta a mesma consistência, rigor e complexidade de pesquisas amostrais sobre o mercado de trabalho brasileiro. Serve como subsídio – e não como fonte oficial – aos institutos geradores de estatísticas como o IBGE, DIEESE, FIESP e centrais sindicais. Algumas informações, como o registro de rendimentos, podem apresentar informações mais corretas, mas há claro *trade off* nesses quesitos. Reinhard (1986) *apud* Jannuzzi (*op. cit.*) coloca que a maior vulnerabilidade que existe na RAIS afetaria acentuadamente os dados de educação – o que é amplamente corroborado por De Negri et al. (2001) e Corseuil e Santos (2002) – e na imputação de categorias de CBO, o que seria causado pelo desconhecimento desse índice arcano de profissões por parte dos informantes. É possível especular sobre a natureza desses desvios: a coleta de informações por heteroatribuição provavelmente afeta as respostas, uma vez que há incentivos para empregados mentirem sobre seus currículos. Além disso, as empresas declarantes podem limitar-se a declarar somente a escolaridade requerida pelo vínculo empregatício, deixando de fora a escolaridade real do indivíduo (CORSEUIL & SANTOS, 2002).

⁸ Admite-se, entretanto, que há perda de qualidade das informações em municípios menores e em alguns setores econômicos, como consta da Nota Técnica MTE 093/2014: “A omissão é frequente em municípios menores. Em alguns setores, percebem-se informações qualitativamente mais comprometidas que em outros. Insere-se nesse contexto a Agricultura, a Administração Pública e a Construção Civil. Nesse sentido, o MTE tem alertado para a leitura das Notas Técnicas e dos Comunicados que são divulgados, frequentemente, contendo informações a respeito de situações que fogem ao padrão normal”. (MTE, 2014, p.5)

⁹ Esta lei compreende os artigos 352 a 371 da CLT. Versa que ao menos dois terços dos empregados de uma empresa devem ser nacionais, exceto para regulação específica.

Como dito, uma comparação entre as duas bases de dados só é possível para o mercado de trabalho formal, incluindo-se o funcionalismo público brasileiro e excluindo algumas categorias, principalmente empregados domésticos com carteira assinada. Contudo, para as observações de cunho racial, o poder público não se manifesta sobre a variável. É importante salientar que, até o momento da publicação dessa monografia, não há instrumentos eficientes de acompanhamento do perfil racial do setor público brasileiro, ou mesmo restrito a União, posto que a qualidade das informações de Cor ou Raça no SIAPE era deveras deficiente. No momento, há propostas para a obrigatoriedade da declaração dessa variável para o serviço público na forma do Projeto de Lei 7.720/2010¹⁰, contudo sem aprovação provável no horizonte próximo. É alarmante que, na esteira da proliferação de leis como a 6.738/2013 (e muitas outras em esfera municipal e estadual), não existam ainda mecanismos de acompanhamento dos resultados concretos desses dispositivos legais. Para tais casos, o acompanhamento das políticas públicas não pode ser feito nem pela PNAD, nem pelo Censo. Isto se dá, respectivamente, por deficiências no tamanho da amostra e de questionário.

II.2 – PERÍODO DE REFERÊNCIA

Para estimar-se o estoque de trabalhadores empregados formalmente em todo o país, o período de referência da RAIS é o do último dia do ano – 31 de dezembro. Em qualquer comparação com pesquisas amostrais com base em um período de referência – para o Censo, é o mês de Julho –, é necessário ancorar-se nesta data para evitar dupla contagem. Isto porque a base da RAIS é composta por vínculos empregatícios incessantemente extinguidos, recriados e substituídos na dinâmica de consistente alta rotatividade do mercado de trabalho brasileiro. Neste sentido, a diferença no período de referência pode causar, na comparação, variações associadas a fenômenos sazonais do mercado de trabalho. Presume-se, tal efeito sazonal é observável através de valores de salários, de nível de emprego e horas trabalhadas maiores, enfatizando-se o efeito sobre a agropecuária, determinados setores de serviços e comércio.

¹⁰ Mesmo aprovada esta lei, a coleta efetiva dos referidos dados depende de uma disciplina administrativa e orientação (senão sanções) aos funcionários públicos e corpos administrativos para que preencham os dados com veracidade. “Leis, leis, leis. Só faltou aquela a que se referiu Ferreira Vianna: a lei que mandaria pôr em execução todas as outras” (PRADO, 1928).

Corseuil e Santos (2002) testaram estas hipóteses utilizando os dados da PME de 1998. Constataram que o efeito sazonal de dezembro em relação a setembro pode ser, de fato, uma explicação para parte dos diferenciais encontrados. A proporção de trabalhadores com baixa carga horária (menor que 12 horas semanais) aumenta de maneira sensível, além de orientar levemente a composição das faixas salariais para níveis mais concentrados nas faixas superiores. A renda média de dezembro é cerca de 10% a 15% maior em comparação aos demais meses; porém, a maior parte deste aumento se concentra na população analfabeta ou que possui entre 8 e 11 anos de estudo, ao passo que nos demais níveis educacionais os ganhos são inexistentes. Estes efeitos são, em tese, observados para vagas temporárias criadas no fim do ano, e não devem afetar a estimativa de renda média anual existente em cerca de 80% do estoque de trabalhadores do setor privado em 31/12 na RAIS de 2010. Contudo, afetam aquelas observações referentes a vínculos empregatícios criados no período que vai de agosto até dezembro. De Negri et al. (2001) testaram a opção de remover da amostra os vínculos empregatícios firmados entre outubro e dezembro do ano de 1998, como forma de limitar o estoque de dezembro àqueles trabalhadores existentes empregados em setembro do mesmo ano. O procedimento resultou em um leve aumento salarial médio, de 5,13 para 5,33 salários mínimos - doravante s.m. - enquanto que levou a uma redução considerável no número de observações: de aproximadamente 35 para 24,5 milhões. Igualmente, visando-se controlar o efeito sazonal da RAIS, com isto a tornando mais compatível aos dados do Censo demográfico, poderia se aventar a remoção, na base da do MTE, dos vínculos firmados a partir de agosto. Contudo, esta alternativa revela-se arbitrária e não parece trazer bons resultados em termos da qualidade das informações que são geradas.

II.3 – RENDA DO TRABALHO

A variável de rendimentos é a que possui maior diferença conceitual entre as duas bases de dados e exige, por este motivo, descrição detalhada das diferenças existentes. As variáveis de renda na RAIS vêm em quatro variedades: renda média do vínculo, renda em dezembro e a transformação linear de ambas em salários mínimos. Para os fins da presente análise, só nos interessa a primeira. Corseuil e Santos (2002) atribuem a maior parte da diferença entre os rendimentos à questão conceitual, destacando que, na PNAD (conceitualmente semelhante ao Censo Demográfico), tais rendimentos são líquidos de tributos e encargos. Os autores relacionam isso ao resultado de os valo-

res brutos dos salários na RAIS serem proporcionalmente superiores em faixas de renda nas quais caberiam alíquotas de imposto de renda mais elevadas. Esta informação é equivocada. Souza (2013) esclarece que o conceito utilizado na PNAD e no Censo é o de renda monetária bruta do trabalho (aquém das recomendações internacionais), uma vez que estas pesquisas não discriminam “rendimentos não monetários, tributos e contribuições” (p. 22). Porém, deve-se ter cuidado. A reportagem de rendimentos obtidos por entrevista é dificultosa e pode facilmente, por virtude de confusão, incluir informações de renda líquida de tributos. Contudo, o questionário é claro e não há razão para se acreditar que uma parte expressiva dos entrevistados tenha declarado a sua renda líquida.

Na verdade, a diferença média pode ser explicada pela introdução de benefícios a serem considerados na imputação dos dados de rendimento. Souza (2013) informa que os Censos não captam uma ampla categoria de rendimentos do trabalho, como: indenizações do 13º salário, hora extra, participação nos lucros, aviso prévio, saques no FGTS, PIS/PASEP etc. Na RAIS, pelo contrário, os rendimentos do trabalhador são reportados, grosso modo, pelo custo às firmas. Assim, incluem-se 27 tipos de remunerações salariais: 13º salário (em campo específico somente), ajuda de custo, gratificações, adicionais por tempo de serviço, comissões, gratificações de férias, aviso-prévio, hora extra, quaisquer outros valores sobre os quais incidam FGTS ou tributos previdenciários e outros. Por outro lado, há 35 tipos de remunerações que não devem ser incluídas, a maior parte delas sendo multas e direitos recebidos em função de demissões ou rescisões contratuais, como benefícios de 40% do FGTS, PIS/PASEP e indenizações diversas (MTE, 2011). Repetimos: não é incluso nos rendimentos médios o décimo terceiro salário. Com a captação dos dados por vínculos, há também dupla contagem dos trabalhadores que possuam, eventualmente, dois ou mais vínculos empregatícios simultâneos. Nestes termos, não é possível determinar o somatório dos pagamentos e da carga horária, contribuindo para a subestimação dos salários médios (PAIXÃO et al., 2012).

Essas diferenças conceituais, no que devem ou não ser incluídos no cálculo dos rendimentos, constituem, além das incertezas associadas às respostas dos entrevistados ou formulários enviados pelas empresas, uma barreira intransponível para a compatibilização. Em mãos dos dados da Pesquisa de Orçamentos Familiares (POF) de 2008, Souza (2013) estima que a diferença salarial para os trabalhadores com carteira assinada

devido a não imputação de benefícios seja de 14%. Já Barros (2007) estima que a diferença de renda do trabalho em razão dos benefícios (sem saques de FGTS e PIS/PASEP) é de 14,7%, dos quais 10% somente em razão do 13º salário e abono de férias. Esses resultados estão desalinhados aos encontrados em Paixão et al. (2012), para quem a diferença de renda entre a PNAD e a RAIS da média total do Brasil no ano de 2009 é de 16,4%. Valor similar é observado nos resultados para grupos de idade e sexo em Corseuil e Santos (2002). Em relação ao Censo de 2010, excluindo-se os casos de renda ignorada, a diferença é de somente 9,2% (com correção monetária pelo INPC, cerca de 5,8%) – o que parece ser coerente, uma vez que os rendimentos da RAIS não captam o 13º salário.

Diferenças de tratamento nas variáveis de rendimento também são de considerável importância. Na RAIS, os tratamentos são relativamente poucos. Para todas as variáveis, os valores são mantidos, sempre que possível nos seus valores originais, com exceção de *outliers* e registros inconsistentes em relação aos dos três anos anteriores. Os *outliers* e os casos ignorados são corrigidos imputando-se os valores dos vínculos de passados quando possível (MTE, 2014). Nas variáveis de renda, os casos com admissões no ano corrente do registro configuram um caso especial: eles são filtrados conforme a CBO, mês de admissão e de desligamento. Assumirão valores iguais a zero todos os casos com remuneração inferior a 0,3 salário mínimo ou que transgredirem o limiar máximo imputado a cada grupo do filtro (DE NEGRI et al., 2001). Em função disto, o estoque de trabalhadores celetistas em 31/12 de 2010 possui 568.769 registros com valor zero. O valor do rendimento destes vínculos deve ser ignorado, uma vez que não há possibilidade deles não serem remunerados na pesquisa.

Comparando os dados dos Censos Demográficos com os da PNAD, Souza (2013) destaca que a distribuição da variável rendimento é levemente diferente em uma pesquisa e outra. Por conta do tamanho da amostra – cerca de 50 vezes maior no Censo –, a probabilidade de captura de rendimentos muito grandes é desproporcionalmente maior do que na PNAD. Por essa razão, a distribuição dos rendimentos, no Censo, é log-normal apenas até certo ponto, havendo um descolamento da distribuição no topo: ela torna-se a partir daí uma distribuição de Pareto. Embora representem uma parcela pequena da população, estes valores desproporcionais podem modificar consideravelmente os resultados de indicadores estatísticos convencionais (SOUZA, 2013). Na

comparação com a RAIS, isto não se manifesta como um problema, uma vez que compartilha a virtude de grandes números. De fato, o valor dos rendimentos nos quantis de rendimento 99º e 999º - nos quais se pode observar a mudança no tipo de distribuição - são um pouco maiores que os do CENSO demográfico, refletindo, em parte, os benefícios que constituem porção dos rendimentos na RAIS. Portanto, as duas bases compartilham o mesmo distúrbio na cauda superior da distribuição da variável log-salário, com efeitos semelhantes sobre indicadores estatísticos. Este fato é de especial relevância no contexto de pesquisas de desigualdade de renda; espera-se que por meio dele se revelem resultados mais aproximados da realidade do mercado de trabalho.

Os desvios mais sérios que abatem as variáveis de rendimento no Censo de 2010 são os erros de coleta. Para ajustar estes defeitos, o IBGE, no processo de crítica dos dados, empreendeu uma série de ajustes naqueles valores considerados anormais (atribuídos a erro de imputação) ou que se encontravam fora da curva. Para isto foram desenvolvidos métodos para detecção destes casos. Criaram-se escores com base nas propriedades do domicílio, de forma a confrontar as qualidades do recinto com a renda declarada pelos entrevistados. Os casos anômalos foram então imputados com valores de doadores aleatórios por meio de *hot deck* com base em um filtro de uma série de variáveis secundárias (IBGE, 2012; 2013).

Porém, estes procedimentos não foram realizados no ano de 2010 para os casos em que os valores declarados foram iguais a zero. De acordo com o IBGE (2012), um total de 4,3% dos domicílios do universo apresentou rendimento com valor zero. Este número é semelhante ao do somatório de casos com valor zero e ignorados na PNAD de 2009. De acordo com o Instituto, isto pode ser atribuído às dificuldades com o sistema de *PDA* utilizado nas entrevistas, o que teria levado os usuários a imputar casos ignorados como possuindo renda zero. De fato, a análise dos dados revela um perfil altamente inconsistente desse grupo de renda: muitos desses casos seriam compostos por grupos ocupacionais de status elevado. Apesar das tentativas bem sucedidas de corrigir esse defeito através de um modelo logístico, uma dificuldade de caráter técnico impediu a correção dos dados em questão, dos quais se chegou à conclusão de que continham 47,8% de casos ignorados. A causa de tal dificuldade, de acordo com o IBGE, seria que:

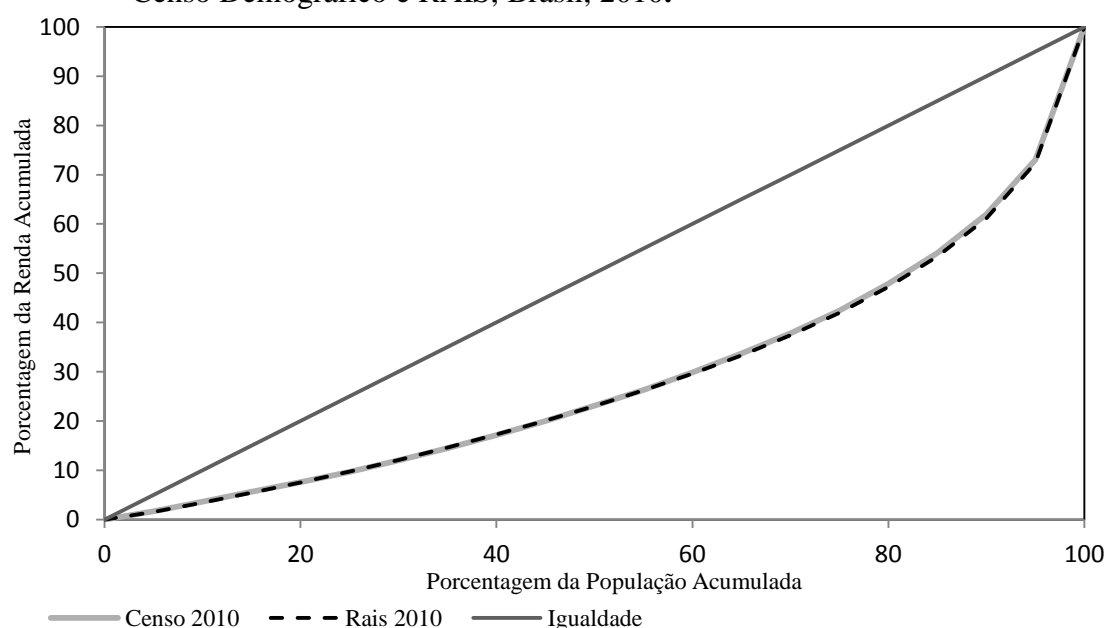
Ao verificar nos dados da Amostra quem eram as pessoas residentes nos domicílios apontados para a imputação constatou-se que, em aproximadamente 90,0% dos casos, eram

pessoas não-ocupadas. Isso implicaria em uma das duas seguintes possibilidades: 1 - Imputar toda a parte de trabalho e rendimento destas pessoas; ou 2 – Imputar somente um rendimento de outras fontes para estas pessoas. A opção mais factível seria a imputação apenas de um rendimento de outras fontes. A distribuição do rendimento de outras fontes é naturalmente diferente da distribuição dos rendimentos de trabalho. Como não há esta separação nos dados do Universo, as distribuições de rendimento nos domicílios imputados da Amostra e do Universo seriam consideravelmente diferentes. Dadas estas restrições, optou-se por não realizar a imputação de rendimento nestes domicílios e manter o valor de rendimento zero (IBGE, 2012, p. 17-8).

Para o grupo estudado, os valores iguais a zero não são uma dificuldade para a comparação no mercado de trabalho; o são apenas para estudos mais abrangentes com enfoque nos domicílios. Embora os casos de rendimento do trabalho com valor igual a zero representem 4,1% do total da PEA ocupada, os mesmos não se encontram na PEA com carteira assinada na semana de referência (o que reforça o realismo destes dados).

Apesar de haver diferenças nos valores dos rendimentos médios do trabalho entre as duas pesquisas – diferenças estas que podemos atribuir à dessemelhança dos conceitos empregados –, os resultados obtidos para o Índice de Desigualdade de Gini são surpreendentes. O valor estimado é de 0,4374 para a RAIS e de 0,4328 para o Censo de 2010. Esta semelhança parece constituir um índice eloquente que permite conjecturar que, de fato, as duas pesquisas são bastante comparáveis estruturalmente. Abaixo, a representação da Curva de Lorenz para estes Índices que, conforme é possível notar se revelam praticamente idênticos:

Gráfico 1 - Curva de Lorenz para a Renda do Trabalho/Vínculo do Censo Demográfico e RAIS, Brasil, 2010.



Fonte: Elaboração própria. Microdados Censo Demográfico 2010, IBGE; Relação Anual de Informações Sociais 2010, MTE.

II.4 – VARIÁVEL DE COR OU RAÇA

A variável Cor ou Raça foi introduzida nas bases de dados do MTE através da portaria nº 1.740/1999 e implementada no ano seguinte. Até recentemente, seu acesso era restrito apenas aos usuários autorizados; contudo, atualmente os dados referentes aos anos de 2007 a 2013 são de domínio público. Embora a portaria citada oriente, explicitamente, o uso da categoria de raça/cor conforme a classificação utilizada pelo IBGE e, implicitamente, a ordem dos questionários utilizados pelo instituto, esses critérios não foram adotados nas pesquisas do MTE. A título de esclarecimento, a ordenação da categoria em questão empregada pelo IBGE é:

- I – branca;
- II – preta;
- III – amarela;
- IV – parda, e
- V – indígena (MTE, 1999).

Do que consta nos questionários, as categorias disponíveis aos usuários são as do IBGE, dispostas em ordem diferente e com um acréscimo no final:

- 1. Indígena – para a pessoa que se enquadrar como indígena ou índia.
- 2. Branca – para a pessoa que se enquadrar como branca.
- 4. Preta – para a pessoa que se enquadrar como preta.
- 6. Amarela – para a pessoa que se enquadrar como de raça amarela (de origem japonesa, chinesa, coreana, etc.).
- 8. Parda – para a pessoa que se enquadrar como parda ou se declarar como mulata, cabocla, cafuza, mameluca ou mestiça de preto com pessoa de outra cor ou raça.
- 9. Não informado (MTE, 2011).

Não é possível avaliar as diferenças causadas pela ordem diferente dos questionários, mas é provável que não sejam negligenciáveis. Como destaca Paixão et al. (2012), nem mesmo o padrão parece ter um critério lógico. O acréscimo da categoria “não informado” também é um problema grave, apesar de ser, aparentemente, uma concessão às deficiências na forma de coleta destes registros. Sugere-se uma alteração simples, mas talvez com resultados modestos: a correção da ordem do quesito posteriormente.

Contudo, a maior diferença que há entre as duas pesquisas é aquela que pode ser atribuída ao informante. Nas pesquisas domiciliares, são entrevistados os seus membros; na ausência de um deles, um dos presentes relata as informações do ausente e o mesmo ocorre para indivíduos incapazes de oferecer respostas. Embora nestes processos

de entrevista haja uma grande mistura entre informações primárias e de terceiros, elas são proferidas por membros do domicílio, ou seja, indivíduos que compartilham a mesma cultura domiciliar e que, portanto, podem oferecer informações relativamente precisas sobre os ausentes (OSÓRIO, 2003). Assim, no Censo Demográfico, como em tantas outras pesquisas, o método de declaração de Cor ou Raça é o da autoatribuição – método este mais recomendado para questionários sobre o tema.

Na RAIS, pelo contrário, o método é heteroatributivo – isto é, a prestação das informações é feita por terceiros. Não há como generalizar a experiência de declaração dos usuários do sistema RAIS. É certo que é feita, amiúde, na ausência de prestação de informações pelos próprios empregados, variando sensivelmente entre as firmas declarantes. Mesmo sendo recomendada a declaração por parte das empresas, esses dados são obtidos com frequência pela matriz, impondo severo grau de separação entre o informante e seu objeto.

Há um conjunto grande de trabalhos sobre o tema dos dois métodos – autoatribuição e heteroatribuição –, ressaltando-se que, neste caso, os informantes das observações da RAIS não possuem (presume-se) treinamento formal¹¹ para a condução das entrevistas com os funcionários – isto, se elas de fato ocorrerem. Assim como há um grande corpo de trabalhos acadêmicos sobre o tema, não há consenso sobre os benefícios de um método ou outro. Contudo, há alguns resultados concretos observados em diversas pesquisas comparativas entre autoatribuição e heteroatribuição que merecem ser considerados. Osório (2003), ao fazer um resumo da discussão, destaca que os resultados obtidos por heteroatribuição causam, de forma consistente, um desvio para o “embranquecimento” do perfil fenotípico-racial da população pesquisada (ao invés de um “empatecimento”). Parte considerável dos casos era contestada: implicavam um conflito de ideias entre entrevistadores e entrevistados, com os primeiros atribuindo a categoria “branca” para ao menos 70% desses casos discordantes. Sobre as causas de tal fenômeno, Osório (2003) explora certos traços do comportamento presumido dos entrevistadores para além das honestas dificuldades em categorizar casos limítrofes:

¹¹ É importante lembrar que o treinamento de recenseadores é geralmente insuficiente. São necessários muitos temporários e novatos para a condução do recenseamento, afetando a qualidade e minúcia das entrevistas. Serão eles melhores que os burocratas das empresas que declaram RAIS?

Considerando que mencionar a cor das pessoas pode ser visto como uma atitude pouco polida, o embranquecimento poderia ser interpretado como uma “concessão” dos entrevistadores aos entrevistados: se quanto mais preto pior, ver o preto como pardo e o pardo como branco torna-se uma “gentileza” à luz da ideologia racial (OSÓRIO, 2003, p. 17).

Ironicamente, neste sentido talvez as próprias entrevistas dos Censos Demográficos sejam muito semelhantes aos procedimentos heteroatributivos. Relatam Piza e Rosemberg (1999) acerca da experiência de entrevistadores durante o Censo de 1991:

Na entrevista nº1, o pesquisador relata que, embora o procedimento estipulado pelo IBGE durante o treinamento tenha sido o de perguntar a cor da população e registrar a resposta da cor auto-atribuída, o pesquisador atribuiu a cor quando: 1) o entrevistado lhe pareceu ser branco; 2) a etiqueta de relacionamento exigiu (como procedimento mais educado) que ele não perguntasse a cor; 3) a automação da entrevista e o cansaço acabavam por abreviar o procedimento de coleta – “[...] no décimo questionário do dia você já não perguntava mais”; 4) só perguntava a cor se o entrevistado fosse negro (significando de cor preta) – “Mas, quando era negro, eu sempre perguntava, porque na instrução tinha muito esse cuidado de enfatizar: Olha, se o cara for negro e falar que é branco, você tinha que pôr” (PIZA & ROSEMBERG, 1999, p. 132).

Não é difícil concluir que os dois procedimentos são enviesados pelas dificuldades práticas do registro dessas informações. É certo que, apesar do declarante/declarado, os dados contêm em grande monta os hábitos e preconceitos dos entrevistadores – mais que dos entrevistados.

Para Telles e Lim (1998), os dados obtidos por meio de heteroatribuição de pertença são superiores na medida em que indicam uma profundidade maior de desigualdades sociais entre os grupos de cor ou raça. Pois, segundo os autores, a discriminação racial de outrem é, no âmbito do mercado de trabalho, um fator mais poderoso que os efeitos internalizados pelo preconceito racial. Valle Silva (1999) chega a conclusões semelhantes ao refletir sobre o efeito que o nível social do indivíduo tem sobre a resposta de cor ou raça. Na hipótese de a pesquisa ser conduzida por pessoal treinado, seria possível eliminar este viés. Osório (2003), por sua vez, conclui mais sensatamente que, em primeiro lugar, somente com um “conjunto imaginário de observadores perfeitos” (p. 13-4) poder-se-ia dar um julgamento unânime para cada indivíduo pesquisado. E, em segundo lugar, a escolha entre as duas opções reflete somente a escolha entre duas subjetividades – a do sujeito ou a do observador. Não obstante essas considerações, podemos chegar a alguns fatos estilizados importantes para a comparação: em tese, a heteroatribuição realizaria um embranquecimento da população (aglutinando casos limítrofes

entre os fenótipos de cor junto à categoria “branca”) e tenderia a acirrar as desigualdades de renda.

É fato que, independentemente do método utilizado, a discriminação racial é detectada tanto pela autoatribuição quanto pela heteroatribuição. Em oposição a Telles e Lim (1998) e Valle Silva (1999), Osório (2003) atribui um resultado mais racional à autoatribuição. Segue simplificadamente o raciocínio: se é verdade que a brancura é positivamente estimada na ideologia racial brasileira (e o inverso ocorre com a pretidão), então haveria uma tendência que entrevistadores treinados atribuísem um perfil mais negro que no caso da autoatribuição. Entretanto, isto não apenas não acontece como a divergência é, aparentemente, tão mais grave quanto melhor o treinamento. Logo, há uma veracidade maior nos dados obtidos por autodeclaração, já que implicam em uma avaliação mais negativa de si. Contudo, podemos argumentar que essas considerações presumem um comportamento racional (utilitário) para o que pode não o ser na prática. Desta maneira, até mesmo essas considerações podem ser contestadas.

Alguns dos resultados de estudos mais recentes parecem indicar a persistência destes fenômenos. É particularmente interessante o caso da pesquisa elaborada pelo IBGE “Características étnico-raciais da população” (PCERP, 2008). Chama a atenção o fato de que os entrevistadores “não receberam nenhum tipo de instrução nem foram orientados sobre como preencher este quesito (classificação de cor ou raça)”, quando suscitados a atribuir classificação de cor ou raça aos entrevistados (IBGE, 2011, p. 31). Assim, embora procedesse a pesquisar a cor ou raça do entrevistado conforme categorias abertas¹² permite uma visão aproximada dos desvios contidos entre os dois métodos de entrevista, sem desvios associados ao treinamento especializado de entrevistadores. Para todas as regiões pesquisadas há uma forte tendência à heteroclassificação¹³ atribuir um perfil populacional mais branco. Por exemplo, no caso do Rio Grande do Sul, há uma diferença de 12,5 pontos percentuais na população branca. A concordância entre entrevistadores e entrevistados é de 92,8% para esse Estado, e 90,6% para o Brasil. Contudo, é significativamente menor nas demais classificações em função da multiplicidade de categorias que, presume-se, indicam população parda (para o qual há concor-

¹² Fora as categorias habitualmente utilizadas pelo IBGE, há também respostas com as classificações “morena”, “negra” e “outras”. Conferir tabelas 2.9 e 2.10 (p. 42-43).

¹³ Conforme denominação utilizada na pesquisa citada acima.

dância de somente 54,2% para o país). Estes resultados são significativos, pois, como será visto, não se confirmam para todas as Unidades Federativas na RAIS.

Finalmente, devemos destacar os resultados de outras duas pesquisas significativas realizadas com bases de dados originais sobre o tema: a primeira é a PESB 2002, sobre a qual Bailey, Loveman e Muniz (2013) fazem uma análise estatística original; a segunda é do PERLA¹⁴, com dados inéditos obtidos através de entrevista com uso assistencial de paletas de cores para identificação dos grupos de cor ou raça. Estes dados foram discutidos por Silva e Paixão (2014) no livro *Pigmentocracies*.

Bailey, Loveman e Muniz (2013) se preocupam em esclarecer – ao testar as várias formas classificatórias –, que a opção de classificação (incluindo formas binárias de identificação) empregada é apenas uma das dimensões de um constructo multifacetado/dimensional que é a característica étnico-racial de um indivíduo. O propósito explícito de identificar qual é a forma mais precisa significa incorrer em um desvio ou erro de percurso – desautorizando em parte o propósito desse trabalho. Advogam simplesmente o uso frequente de formas classificatórias das mais diversas possíveis (contudo negligenciando as dificuldades técnicas e políticas disso). Na análise de regressão quantil-simultânea (KOENKER & BASSET, 1978) utilizada, encontram que no formato censitário e heteroatributivo, para todos os decis de renda média horária a relação entre os salários da população preta para com a branca é inferior aos encontrados por métodos autoatributivos, mas não é o caso da parda.

No caso do uso de um modelo binário – isto é, que coteja somente as categorias preta e branca –, verifica-se uma maior desigualdade de renda em todos os casos quando o método empregado é o de heteroatribuição. Isto sugere que a hipótese de Telles e Lim (1998) é verdadeira, ou seja, de que a percepção direta de terceiros é fundamental na identificação racial dos indivíduos, e fonte substancial de discriminação no mercado de trabalho (BAILEY, LOVEMAN & MUNIZ, 2013). Por fim, os autores identificam haver de fato uma preferência pela identificação branca para a população parda, quando os entrevistados são forçados a escolher apenas por categorias binárias – reforçando assim algumas das teorias de ideologia racial discutidas. Porém, vale ressaltar que a PESB 2002 difere significativamente das demais pesquisas em uma característica: não se ob-

¹⁴ Acrônimo para *The Project on Ethnicity and Race in Latin America*, Universidade de Princeton.

serva, nesta base de dados, o processo de embranquecimento. Na verdade, enquanto que no formato censitário por autoatribuição, 52,2% da amostra se declara branca, por heteroatribuição este valor é de 49,1%.

Os dados do PERLA, esmiuçados por Silva e Paixão (2014), dos trabalhos citados nesta seção, apresentam a maior concordância entre os métodos classificatórios conforme as categorias do IBGE: em torno de 84% para pretos, pardos e brancos. Similarmente, as metodologias de classificação baseadas no matiz de cor de pele e outros traços fenotípicos mantêm uma forte relação entre essas categorias censitárias. Mesmo havendo grande concordância entre os diferentes métodos heteroatributivos e autoatributivos, quando os mesmos são comparados as médias de anos de estudo – que podem ser utilizadas como *proxy* de renda – os primeiros mostraram mais fortemente uma relação linear entre a presença de fenótipos brancos e uma média superior para o indicador. O inverso ocorre para fenótipos mais escuros. Estes resultados, assim como os apresentados por Bailey, Loveman e Muniz (2013), parecem concordar novamente com Telles e Lim (1998). Porém, uma vez que só se tem em mãos os indicadores de média de anos de estudo, estas observações não constituem resultado forte e, conforme admitem os autores, não são suficientes para aplacar as pautas de pesquisa que não atribuem centralidade ao papel da discriminação racial nas desigualdades de renda (SILVA & PAIXÃO 2014).

Quais fatos estilizados importantes podem ser derivados desta revisão sobre o tema da heteroatribuição e autoatribuição? Mais uma vez, não há consenso. O resultado mais forte a ser esperado é a tendência ao embranquecimento, isto é, a uma composição maior de população branca pelo método de heteroatribuição. O segundo e mais contencioso resultado é a possibilidade de detectar maior discriminação racial de renda no mercado de trabalho. A observação simultânea (ou não) destes dois fatos pode servir como importante ponto de partida para a produção de mais trabalhos sobre o tema – em especial para aqueles de cunho sociológico; o teste para teorias como o da “válvula de escape do mulato” (DEGLER, 1971), a importância de aspectos sociais e de classe mais básicos sobre discriminação racial etc. Não é o propósito deste trabalho a análise dessa discussão, porém os fatos aludidos serão testados.

II.5 – OUTRAS CONSIDERAÇÕES

Embora os modelos econométricos favoreçam o uso de renda média horária, não é possível executá-los de tal modo: na RAIS não são imputadas as horas efetivamente trabalhadas, mas sim as horas contratadas; já no Censo, é justamente o contrário. Desta forma, são subestimadas na RAIS em comparação ao Censo e podem, assim, provocar desvios ainda maiores na comparação da renda auferida.

As variáveis de educação podem ser livremente manipuladas no Censo de 2010 e não representam problema de compatibilização com a RAIS. Mas, como já foi dito, há diferenças significativas associadas com as divergências de método de atribuição, mesmo que formalmente os grupos formados sejam idênticos. Os resultados empíricos constatarem essa diferença na composição dos grupos educacionais, podendo, neste sentido, provocar dificuldades em modelos baseados na teoria do capital humano.

Recapitulando: as compatibilizações podem ser feitas com um mínimo de interferência, se restringem ao grupo estudado e à eliminação de casos de renda ignorada ou igual a zero. Optou-se por manter no estudo os casos de cor não informada, já que: (1) estão levemente acima do nível de 5%; e (2) não há certeza se a omissão dos casos é um fenômeno aleatório ou não. Consideramos que eles constituem, por si só, um fenômeno importante a ser estudado. É perfeitamente possível imputar o valor de doadores a casos sem informações. Por limitação de *software*, ficamos restritos à imputação através de um modelo de regressão linear, o que entra em conflito com os propósitos do trabalho. O mais desejável seria um filtro especializado que permitisse a imputação de doadores aleatórios dentro dos grupos para a eventualidade desses erros não serem aleatórios (*hot deck*).

As análises que seguem partirão de um sistema binário de cor, isto é, serão utilizados somente os grupos “brancos” e “pretos e pardos”¹⁵. Isto se dá a partir do entendimento compartilhado de Valle Silva (1980) de que não há diferenças significativas entre os componentes do último grupo dentro do mercado de trabalho.

¹⁵ Contudo não se deve desprezar a análise de modelos trinários. Por exemplo, Bailey, Loveman e Muniz (2013) encontram diferenças nos rendimentos das duas populações, isto é a de pretos e pardos conforme a metodologia utilizada. A opção por um modelo binário reflete a sua maior facilidade de uso e espera-se - considerando que as diferenças entre pretos e pardos são menores que desses em relação aos de brancos - que a adoção de um modelo e outro não afete significativamente os resultados encontrados. Daí a opção, que se reconhece sempre passível de crítica, de seguir Valle Silva (1980) e utilizar o modelo binário.

CAPÍTULO III – DISCUSSÃO DOS DADOS QUANTITATIVOS

Com o auxílio de tabelas, o presente capítulo trata de comparar, preliminarmente, os dados da RAIS e do CENSO de 2010, sendo dividido por duas seções: a primeira para a discussão dos dados por Unidades da Federação e a segunda conforme o nível educacional da população.

III.1 – DADOS DAS UNIDADES FEDERATIVAS.

Na tabela abaixo¹⁶, encontram-se dados sobre a composição da população empregada com carteira assinada nas Unidades Federativas e da distribuição conforme sexo e cor ou raça. As colunas com os dados das UFs e de gênero permitem avistar como são retratadas, nas duas fontes de dados – com as divergências metodológicas discutidas –, perguntas simples e definitivamente objetivas, sobre as quais não pesa as dificuldades sociológicas mencionadas. Como pode ser visto, não há diferenças significativas na composição da população dos Estados. Em termos relativos, alguns dos menores apresentam certa incongruência. O mesmo ocorre no caso de São Paulo, que possui 30,07% da população ocupada com carteira segundo o Censo, mas 32,31% segundo a RAIS em 2010. Isso se dá possivelmente em função de pequenos erros associados à expansão da amostra, além dos efeitos da subdeclaração na RAIS. A diferença constatada entre a distribuição da população de sexo feminino ocupada é pequena, de 1,04 ponto percentual. A diferença é maior no Nordeste, sensível em 2,42 pontos percentuais.

A coluna com a distribuição dos casos de cor ou raça não declarados na RAIS mostra decididamente um comportamento não aleatório. Este valor varia entre 3,38% no Rio de Janeiro e até 17,65% no Acre. Estados pouco desenvolvidos (isto é, com PIB per capita inferior), como Roraima, podem apresentar um total de 10,74% de casos ignorados; similantemente, Santa Catarina, com maior nível de desenvolvimento, exibe proporção ainda maior: 11,35%. A partir disso, pode-se concluir que não há relação direta clara entre uma maior cobertura e facilidade de acesso sobre a proporção de casos não declarados. O total destes últimos é de 5,14%. De fato, no curso da presente pesquisa, se identificou que a variável de maior correlação com os casos não declarados é justamente o de região geográfica (seção IV.1).

¹⁶ Não seremos detidos por descrições minuciosas dos dados; apenas pelas tendências mais significativas, face aos fatos estilizados que procuramos explicar. Para tanto se inclui as tabelas com seus valores integrais.

Tabela 1

Composição por UF e distribuição por Sexo e Cor ou Raça da população ocupada com carteira assinada por UFs. Comparação entre Censo e RAIS, 2010.

CENSO		RAIS	Distribuição	CENSO*			RAIS				Diferença	
Composição				Fem (%)	Branco (a)	Preto e Pardo (b)	Fem (%)	Branco (c)	Preto e Pardo (d)	Cor ou raça Não Decl.	(a-c)	(b-d)
100	100	BRASIL	38,58	53,58	45,19	37,54	59,71	34,19	5,14	-6,13	11,00	
4,69	4,31	Norte	34,64	27,64	70,68	32,93	23,70	68,86	5,88	3,94	1,82	
0,68	0,64	Rondônia	37,00	35,42	62,54	33,40	33,33	56,63	7,87	2,09	5,91	
0,26	0,19	Acre	41,07	26,45	71,27	37,35	17,83	63,25	17,65	8,62	8,02	
1,13	1,08	Amazonas	36,98	26,90	71,40	35,73	20,78	73,49	3,82	6,12	-2,09	
0,12	0,11	Roraima	42,17	25,47	70,79	40,14	24,13	63,91	10,74	1,35	6,88	
1,86	1,76	Pará	31,20	25,74	73,08	30,04	22,75	72,35	3,78	2,98	0,73	
0,22	0,17	Amapá	38,57	27,47	71,11	38,22	17,48	73,07	8,68	10,00	-1,96	
0,42	0,35	Tocantins	31,56	27,06	70,58	30,72	25,97	62,17	9,63	1,09	8,40	
16,90	15,80	Nordeste	36,21	31,58	66,90	33,79	27,96	64,18	6,53	3,61	2,72	
1,33	1,10	Maranhão	31,65	27,69	70,84	30,53	24,26	69,68	4,40	3,43	1,16	
0,80	0,67	Piauí	34,95	27,30	70,47	34,65	18,76	73,54	6,26	8,53	-3,06	
2,81	2,72	Ceará	39,10	34,80	63,69	37,84	27,66	66,50	4,28	7,14	-2,81	
1,22	1,08	Rio Grande do Norte	38,23	43,33	55,49	34,57	35,79	47,45	15,47	7,54	8,04	
1,03	0,92	Paraíba	34,73	41,78	56,46	30,76	43,48	49,66	5,10	-1,70	6,79	
3,12	3,29	Pernambuco	35,29	38,69	59,99	33,10	39,27	53,14	6,20	-0,57	6,86	
0,96	0,92	Alagoas	33,06	33,73	64,80	26,19	24,71	67,04	6,82	9,02	-2,24	
0,74	0,71	Sergipe	35,92	28,84	69,29	33,61	28,44	65,68	4,10	0,40	3,61	
4,90	4,39	Bahia	37,05	21,87	76,57	34,54	17,43	74,55	7,19	4,45	2,01	
52,41	54,26	Sudeste	38,92	56,12	42,71	38,22	65,39	29,93	3,81	-9,28	12,78	
10,86	10,63	Minas Gerais	37,03	45,62	53,24	36,68	56,48	38,66	3,90	-10,86	14,58	
1,99	1,95	Espírito Santo	37,69	40,17	58,82	35,95	43,90	51,48	3,54	-3,73	7,34	
9,48	9,37	Rio de Janeiro	38,62	48,29	50,83	37,19	55,97	39,75	3,38	-7,69	11,07	
30,07	32,31	São Paulo	39,78	63,43	35,28	39,17	72,35	22,90	3,93	-8,92	12,38	
18,30	18,26	Sul	41,72	78,44	20,81	40,74	83,63	8,51	7,33	-5,19	12,30	
6,88	6,64	Paraná	40,98	69,90	28,97	39,14	80,46	12,39	6,46	-10,56	16,58	
4,82	4,92	Santa Catarina	42,61	83,79	15,62	41,86	83,44	4,65	11,35	0,34	10,98	
6,60	6,70	Rio Grande do Sul	41,83	83,43	16,10	41,50	86,91	7,51	5,24	-3,48	8,59	
7,70	7,38	Centro-Oeste	36,36	41,35	56,73	35,40	47,71	44,54	6,00	-6,35	12,19	
1,27	1,19	Mato Grosso do Sul	35,21	46,79	50,78	34,45	53,20	37,51	7,49	-6,41	13,27	
1,56	1,48	Mato Grosso	32,85	37,80	60,65	31,95	37,51	53,74	7,48	0,29	6,91	
3,21	2,81	Goiás	35,03	40,99	57,15	35,53	46,76	45,04	6,11	-5,77	12,11	
1,66	1,90	Distrito Federal	43,14	41,23	56,77	38,48	53,64	41,01	3,77	-12,42	15,77	

Fonte: Elaboração própria. Microdados Censo Demográfico 2010, IBGE; Relação Anual de Informações Sociais 2010, MTE.

Nota: * Com expansão da Amostra

Para o Brasil, de acordo com o Censo, 53,58% da população ocupada com carteira assinada se classificou como sendo de cor ou raça branca, enquanto que 45,19% se declarou preta ou parda. Já na RAIS, a 59,71% da população foi atribuída cor ou raça branca, mas apenas 34,19% preta ou parda. Há uma diferença de 6,13 pontos percentuais para a distribuição da população branca, enquanto que este valor é de 11 pontos para pretos ou pardos. Não é possível explicar a diferença nestes dois últimos valores somente em função dos casos não declarados, uma vez que as evidências apontam para uma composição de perfil de cor ou raça aproximada ao médio neste último grupo. Assim, estes resultados marcam o que parece ser uma preferência por imputar vínculos como sendo ocupados por trabalhadores brancos; isto é, com embranquecimento estatístico da população.

Contudo, é importante observar que, embora o resultado para o país seja o de embranquecimento, isto não é verdadeiro para todas as regiões. No que pese o efeito dos casos não declarados, que parecem induzir unicamente a uma redução da população preta e parda, na maior parte da região Norte e Nordeste não se observa embranquecimento. Atente para a penúltima coluna “(a-c)”, que representa a diferença na distribuição de brancos nas duas pesquisas: os valores negativos representam casos em que a RAIS superestima os valores desse grupo em relação ao Censo, quando positivos, subestima. Exceto para Pernambuco e Paraíba, são todos subestimados para as duas regiões. Já para as demais, exceto em Mato Grosso e Santa Catarina (em função da grande quantidade de casos não declarados), são todos superestimados. Ao contrário do que sugerem os resultados da PCERP 2008, a preferência ao embranquecimento não é uma regularidade para todas as regiões e estados do Brasil.

Já quanto aos pretos ou pardos, pode-se observar que, conquanto não exista, regularmente, uma tendência ao embranquecimento, há definitivamente para quase todas as regiões uma subestimação desta população. Como se depreende da última coluna, exceto por quatro estados, o valor da distribuição é subestimado, revelando, se não uma preferência pela classificação “branca”, talvez uma aversão à “preta” ou “parda”. Este fenômeno acontece com maior intensidade no Sul, Sudeste e Centro-Oeste. É bastante óbvio que parte do quadro que estamos vendo é obscurecida pela presença dos casos não declarados. Optamos por não representar a distribuição sem a presença deles, mas é evidente que se forem excluídos, obter-se-á resultados que aumentam a intensidade do

processo de embranquecimento¹⁷. Entretanto, se isto for operado, tampouco irá mudar o fato característico de que não se observa integralmente o fenômeno no território nacional (ocorreria em doze estados da federação). Este fenômeno é observável no país e é predominante em função do tamanho desproporcional que o Sudeste e Sul têm, neste caso, no mercado de trabalho formal.

Ora, se a diferença de treinamento de entrevistadores não pesar em demasia, isso significa que muitos dos resultados de pesquisa sobre o tema podem ser enviesados unicamente por conta do desenho amostral utilizado ou pela abrangência geográfica. Igualmente, se isto for ignorado na forma funcional de testes estatísticos mais detalhados, novamente se pode incorrer em erro. O mais simples seria, talvez, o de tomar o todo como representativo de suas partes e, por exemplo, prescrever teoria de existência de toda uma ideologia racial brasileira em cima deste entendimento (quando poderia ter traços marcadamente mais locais).

Encontra-se na página seguinte a Tabela 2, com formato semelhante à anterior, mas desta vez tratando de comparar o rendimento nominal médio nas duas bases de dados. A renda média para o Brasil na RAIS (do mês de dezembro) foi de R\$ 1407,57, enquanto que no Censo o valor é de R\$ 1288,79 para o ano de 2010 (em julho). Essa diferença de renda já foi atribuída à diferença no cálculo da variável de rendimento do trabalho; o que surpreende é que, ao separarmos os dois grupos, o valor médio do rendimento da população branca é bastante semelhante. É até mesmo inferior em algumas regiões para o último grupo citado, mas em todos os estados (exceto Amapá, Roraima e Distrito Federal), a renda de pretos e pardos é superior na RAIS, sendo até 22,7% maior em São Paulo – o de maior peso no país. As regiões Sul e Sudeste, além da Bahia, Amazonas e Rondônia, são as que possuem maior diferença de renda entre as duas bases de dados, acima de 10%, além de um nível no qual podemos atribuir a incongruência dos rendimentos às diferenças metodológicas da pesquisa.

Nas colunas “(e/f)” e “(g/h)”, constam as razões entre a renda da população branca e da preta e parda no Censo e RAIS, respectivamente. Em 2010, para o Brasil, de acordo com o Censo, pretos e pardos receberam 65% da renda média dos brancos, enquanto que na RAIS esta razão é de 71%.

¹⁷ Isto, é, se admitirmos que esses casos não declarados são apenas uma amostra média da população e que, portanto, não alteram a razão entre a distribuição de brancos e pretos e pardos.

Tabela 2

Renda Nominal Média do trabalho por UFs e Cor ou Raça da população ocupada com carteira assinada por UFs. Comparação entre Censo e RAIS, Brasil, 2010.

Região	Distribuição	CENSO*			RAIS			Diferença		
		Branco (e)	Preto e Pardo (f)	Total (i)	Branco (g)	Preto e Pardo (h)	Total (j)	(e/f)	(g/h)	(i/j)
BRASIL		1531,73	989,07	1288,79	1573,25	1117,39	1407,57	0,65	0,71	0,92
Norte		1440,73	1028,59	1144,64	1399,21	1116,83	1185,43	0,71	0,80	0,97
	Rondônia	1419,25	1064,04	1191,55	1276,56	1212,17	1217,13	0,75	0,95	0,98
	Acre	1513,99	1091,92	1200,65	1241,07	1035,09	1115,86	0,72	0,83	1,08
	Amazonas	1590,88	1070,59	1215,18	1668,01	1263,52	1365,06	0,67	0,76	0,89
	Roraima	1455,83	1046,16	1149,03	1345,47	966,79	1055,57	0,72	0,72	1,09
	Pará	1378,54	995,65	1096,16	1351,44	1056,69	1121,97	0,72	0,78	0,98
	Amapá	1324,71	1119,53	1174,67	1454,13	1006,66	1078,90	0,85	0,69	1,09
	Tocantins	1361,57	923,01	1043,74	1291,50	934,73	1025,20	0,68	0,72	1,02
Nordeste		1308,42	891,95	1024,81	1242,66	955,01	1043,49	0,68	0,77	0,98
	Maranhão	1333,66	900,21	1021,47	1359,15	966,08	1055,68	0,67	0,71	0,97
	Piauí	1255,61	870,48	976,12	1165,94	902,14	947,39	0,69	0,77	1,03
	Ceará	1219,73	835,88	970,49	1149,15	826,77	920,00	0,69	0,72	1,05
	Rio Grande do Norte	1223,67	902,14	1042,92	1118,90	924,03	995,56	0,74	0,83	1,05
	Paraíba	1140,49	821,44	955,57	1031,00	831,79	916,60	0,72	0,81	1,04
	Pernambuco	1326,15	870,42	1049,03	1228,20	957,37	1069,76	0,66	0,78	0,98
	Alagoas	1177,04	834,45	949,29	1189,46	872,30	952,84	0,71	0,73	1,00
	Sergipe	1355,03	949,64	1068,13	1194,35	1070,90	1134,64	0,70	0,90	0,94
	Bahia	1510,98	941,28	1067,63	1533,93	1051,20	1154,75	0,62	0,69	0,92
Sudeste		1683,80	1023,90	1407,25	1739,52	1222,04	1589,00	0,61	0,70	0,89
	Minas Gerais	1312,66	908,78	1094,84	1249,52	1060,82	1170,50	0,69	0,85	0,94
	Espírito Santo	1495,29	989,52	1194,00	1336,55	1105,65	1214,47	0,66	0,83	0,98
	Rio de Janeiro	1871,73	1082,12	1466,13	1868,68	1266,23	1653,92	0,58	0,68	0,89
	São Paulo	1743,01	1064,01	1515,66	1851,35	1305,46	1730,73	0,61	0,71	0,88
Sul		1297,89	938,76	1224,54	1323,84	1099,17	1292,93	0,72	0,83	0,95
	Paraná	1331,36	948,41	1225,06	1299,67	1093,88	1268,76	0,71	0,84	0,97
	Santa Catarina	1277,39	957,77	1227,16	1296,00	1113,25	1264,36	0,75	0,86	0,97
	Rio Grande do Sul	1283,69	907,22	1222,08	1365,62	1101,52	1337,84	0,71	0,81	0,91
Centro-Oeste		1592,71	1075,82	1302,60	1442,86	1114,10	1271,63	0,68	0,77	1,02
	Mato Grosso do Sul	1254,19	936,38	1089,24	1187,38	988,48	1097,50	0,75	0,83	0,99
	Mato Grosso	1385,00	994,09	1146,98	1268,94	1061,64	1134,72	0,72	0,84	1,01
	Goiás	1287,72	977,51	1126,90	1224,74	1033,34	1117,44	0,76	0,84	1,01
	Distrito Federal	2654,36	1445,46	1953,22	1979,58	1370,94	1716,99	0,54	0,69	1,14

Fonte: Elaboração própria. Microdados Censo Demográfico 2010, IBGE; Relação Anual de Informações Sociais 2010, MTE.

Nota: * Com expansão da Amostra

Isto denota menor desigualdade de renda observável no mercado de trabalho formal medido pela RAIS, apesar da metodologia de heteroatribuição de pertença. Exceto para os Estados de Amapá e Roraima, a desigualdade racial de renda é marcadamente inferior nos demais. A análise dos dados dessas colunas não parece indicar padrão claro.

A partir da tabela 2, pode-se concluir que um dos fatos estilizados mencionados na seção anterior não se verifica: não há um aumento da desigualdade de renda, pelo contrário. Há muitos cenários de como isso pode ser interpretado tomando o Censo Demográfico como critério de veracidade. É certo (senão provável) haver uma transferência de casos de vínculos de trabalhadores que, em tese, deveriam ser autodeclarados pretos ou pardos para a população branca. De acordo com a tese de Degler (1971) ou a partir de resultados do estudo do PERLA (2014), há de se imaginar que a tendência ao embranquecimento se dá mais fortemente sobre aqueles grupos de indivíduos de fenótipos limítrofes. Portanto, se dá sobre indivíduos que possuem dentro do seu grupo social (pretos ou pardos, ainda assim com status social baixo) uma renda superior ao da média¹⁸, ou até mesmo através de efeito conhecido: “o dinheiro embranquece”. Mas se for o caso, o efeito causado por tal fenômeno seria o de redução simultânea do rendimento médio dos dois grupos – o que não é o observado aqui.

É até mesmo verossímil, embora pouco provável (para uma população suficientemente grande), que haja uma transferência de população de estrato médio branca para a preta e parda, aumentando a renda média deste último grupo e mantendo constante a do primeiro. Se o embranquecimento da população ocupada ocorresse de forma aleatória, então a renda da população branca simplesmente convergiria para a média total, permanecendo imóvel a do outro grupo (presumindo que o deslocamento ocorre unicamente em direção ao primeiro). Também é improvável a tendência de pessoas de baixo status social se embranquecerem, posto que isto não seja uma opção na RAIS – e possivelmente os olhos de terceiros seriam mais inclementes que os próprios; contudo, possuiria os efeitos observados sobre a renda e a distribuição. Há a dificuldade adicional de se explicar a regularidade com que a desigualdade é inferior na RAIS, ao mesmo tempo em que o fenômeno do embranquecimento se dá de maneira irregular conforme a região

¹⁸ Isso se observa também em dados censitários. Embora o inverso possa acontecer, o status social elevado pode induzir a uma tendência do entrevistado de se escurecer. Ver Paixão e Silva (2014).

geográfica. Em suma, manifesta-se aqui a inadequação de certos modelos explicativos. É provável que a solução se encontre em uma multiplicidade de movimentos em quantidades variadas (ou mesmo na inadequação da RAIS para a variável que está sendo investigada).

III.2 – DADOS POR NÍVEL EDUCACIONAL.

Na página abaixo encontra-se a tabela 3. Nela é possível observar a composição do maior nível educacional alcançado na população aqui estudada por sexo e cor ou raça. A primeira diferença a ser notada é o menor analfabetismo na RAIS. Enquanto que atinge somente 0,56% dos casos nesta relação, o Censo indica que 2,34% da população empregada no mercado de trabalho formal se declara analfabeta. Igualmente, nos demais grupos os dados não coincidem: a população com nível superior é de 14,54% do total na última pesquisa citada, mas de apenas 11,30% na RAIS. Os grupos que compreendem os níveis de educação até o fundamental incompleto também são subestimados na RAIS. A maior parte dos casos se concentra nos grupos de ensino médio completo e fundamental completo – os valores são respectivamente de 43,47% e 14,08% na RAIS, enquanto que no Censo Demográfico é de apenas 31,14% e 10,21%. Mais uma vez, pode-se atribuir ao método heteroatributivo as diferenças apresentadas.

Verifica-se uma tendência dos casos se concentrarem no grupo de ensino médio completo em ambas as pesquisas. Fora a constatação de que isto reflete efetivamente o nível educacional da população, essa exacerbação na RAIS se deve, possivelmente, ao fato de que a maior parte dos vínculos empregatícios exige nível médio. Ademais, hipotetiza-se aqui, que o método empregado de pesquisa acaba muitas vezes por responder somente o nível educacional presumido do vínculo, e não do indivíduo. Posto que haja, por exemplo, uma grande população empregada em vínculos de nível médio, mas graduados com nível superior, isto explicaria a subestimação de tal grupo na RAIS. O mesmo vale para os demais que foram subestimados, isto é, dos analfabetos e do grupo com fundamental incompleto.

Apesar dessas diferenças, os dados mantêm relação de ordem, e a razão entre a composição de brancos e de pretos e pardos se mantém relativamente uniforme nas duas bases de dados. Constatadas essas diferenças, pode-se antecipar que elas terão efeito

sensível sobre a mensuração dos retornos monetários associados a tais características do trabalhador¹⁹.

Tabela 3

Composição de maior nível educacional alcançado por Sexo e Cor ou Raça da população ocupada com carteira assinada. Comparação entre Censo e RAIS, Brasil, 2010.

Composição		CENSO			RAIS		
Sexo	Escolaridade	Branco	Preto e Pardo	Total	Branco	Preto e Pardo	Total
Masculino		-	-	-	-	-	-
	Analfabeto	1,78	4,70	3,21	0,46	1,17	0,75
	Até 5º Ano Incompleto	8,37	11,85	10,06	3,61	7,17	4,97
	5º Ano Completo	6,75	8,50	7,59	5,63	6,64	5,96
	6º ao 9º Ano Incompleto	9,51	13,64	11,51	9,58	12,13	10,44
	Fundamental Completo	11,04	11,32	11,15	16,07	15,97	15,93
	Médio Incompleto	8,94	10,71	9,79	9,57	10,16	9,74
	Médio Completo	29,23	29,31	29,23	39,75	40,07	40,19
	Superior Incompleto	8,90	4,93	6,97	4,03	2,25	3,33
	Superior Completo	15,49	5,05	10,49	11,30	4,43	8,67
Feminino		-	-	-	-	-	-
	Analfabeto	0,52	1,60	0,96	0,18	0,31	0,23
	Até 5º Ano Incompleto	4,63	6,46	5,37	1,48	2,28	1,73
	5º Ano Completo	3,78	4,80	4,18	2,96	3,13	2,98
	6º ao 9º Ano Incompleto	5,28	8,06	6,40	5,59	6,78	5,91
	Fundamental Completo	8,37	9,25	8,70	11,15	10,80	11,00
	Médio Incompleto	7,26	9,45	8,15	7,78	8,51	7,99
	Médio Completo	31,78	37,79	34,18	46,27	54,07	48,91
	Superior Incompleto	12,16	9,49	11,07	6,16	4,45	5,58
	Superior Completo	26,22	13,09	20,99	18,43	9,68	15,67
Total		-	-	-	-	-	-
	Analfabeto	1,25	3,64	2,34	0,35	0,88	0,56
	Até 5º Ano Incompleto	6,80	10,00	8,25	2,76	5,54	3,76
	5º Ano Completo	5,50	7,23	6,28	4,57	5,47	4,84
	6º ao 9º Ano Incompleto	7,73	11,72	9,54	7,98	10,34	8,74
	Fundamental Completo	9,91	10,61	10,21	14,11	14,25	14,08
	Médio Incompleto	8,24	10,28	9,16	8,85	9,61	9,08
	Médio Completo	30,30	32,22	31,14	42,35	44,74	43,47
	Superior Incompleto	10,27	6,49	8,55	4,88	2,98	4,18
	Superior Completo	20,00	7,81	14,54	14,15	6,18	11,30

Fonte: Elaboração própria. Microdados Censo Demográfico 2010, IBGE; Relação Anual de Informações Sociais 2010, MTE.

Nota: * Com expansão da Amostra; Valores vazios representam o valor um.

¹⁹ Por exemplo, em Schultz (1973) – expoente da teoria do capital humano –, os retornos econômicos de educação superior seriam associados a uma produtividade superior determinada por treinamento e proficiência adquiridos. A rigor, se for verdadeiro, então estratos de alta e média produtividade podem ser misturados no mesmo grupo na RAIS. Alternativamente, se o vínculo é determinante na produtividade do trabalho e a educação serve como sinal de distinção para a seleção – como nos modelos de Arrow (1973) e Spence (1973) – faz pouca diferença.

A tabela 4, apresentada abaixo, contém os dados de renda do trabalho para os mesmos níveis educacionais estudados. Observa-se igual tendência geral nas duas fontes. O diferencial de renda é mais severo na população com nível superior, e menor nas demais, especialmente no que se refere aos analfabetos. Contudo, há uma diferença considerável no valor dos rendimentos desses grupos entre as duas bases de dados: para a população com nível superior incompleto, a RAIS superestima esse valor em 30,23% em relação ao Censo em 2010; já para o nível superior completo, é superior em 23,34% (para o total da população, o valor é de 9,2%). Para os grupos com nível médio incompleto ou completo o deslocamento dos valores não é significativo. É importante observar que a superestimação dos valores de renda é semelhante entre os dois grupos de gênero com nível superior incompleto e completo; o mesmo não se pode dizer em relação aos demais níveis educacionais. Resulta disso que a diferença de renda por sexo é superior na RAIS em comparação ao Censo demográfico: na primeira, a diferença de renda por gênero é de 30,8%; na última é de 22,2%. Esse imprevisível resultado também poderá ter efeitos significativos na comparação paramétrica das duas fontes de dados. Já para os grupos de cor ou raça, a desigualdade de renda é menor na RAIS em todos os grupos, exceto para analfabetos e na população com superior incompleto.

Tabela 4

Renda Nominal Média do trabalho do trabalho por maior nível educacional alcançado, Sexo e Cor ou Raça da população ocupada com carteira assinada. Comparação entre Censo e RAIS, Brasil, 2010.

Rendimento Médio Nominal Mensal		CENSO*			RAIS			Razão	
Sexo	Escolaridade	Branco (a)	Preto e Pardo (b)	Total	Branco (c)	Preto e Pardo (d)	Total	(b/a)	(d/c)
Masculino		1697,53	1049,33	1385,76	1755,06	1209,53	1544,18	0,62	0,69
	Analfabeto	699,19	659,44	670,67	835,05	719,50	763,52	0,94	0,86
	Até 5º Ano Incompleto	1010,24	832,88	908,61	1020,20	888,06	949,68	0,82	0,87
	5º Ano Completo	1016,93	860,07	932,00	1122,87	997,51	1065,15	0,85	0,89
	6º ao 9º Ano Incompleto	955,09	827,35	880,95	1111,80	977,19	1051,70	0,87	0,88
	Fundamental Completo	1155,50	960,52	1058,83	1157,47	1017,62	1102,48	0,83	0,88
	Médio Incompleto	998,40	844,55	916,21	1079,46	939,37	1022,16	0,85	0,87
	Médio Completo	1326,94	1064,22	1198,99	1444,43	1209,43	1349,57	0,80	0,84
	Superior Incompleto	1816,32	1483,20	1702,81	2356,61	1884,70	2230,61	0,82	0,80
	Superior Completo	4356,42	2960,76	4039,52	5142,88	3743,64	4870,30	0,68	0,73
Feminino		1303,33	873,87	1134,38	1300,32	933,84	1180,78	0,67	0,72
	Analfabeto	608,60	573,77	585,92	671,45	621,52	648,91	0,94	0,93
	Até 5º Ano Incompleto	806,44	662,00	737,15	718,14	658,01	694,38	0,82	0,92
	5º Ano Completo	743,20	655,91	702,85	743,14	678,74	720,63	0,88	0,91
	6º ao 9º Ano Incompleto	688,74	630,29	658,74	734,46	664,48	708,61	0,92	0,90
	Fundamental Completo	830,76	715,74	781,58	775,53	695,25	750,35	0,86	0,90
	Médio Incompleto	749,34	653,53	704,37	771,05	674,88	737,33	0,87	0,88
	Médio Completo	907,15	758,77	845,43	959,72	819,90	907,73	0,84	0,85
	Superior Incompleto	1164,31	967,30	1097,02	1491,55	1226,89	1419,82	0,83	0,82
	Superior Completo	2458,37	1780,85	2295,60	2926,28	2262,61	2786,25	0,72	0,77
Total		1531,73	989,07	1288,79	1573,25	1117,39	1407,57	0,65	0,71
	Analfabeto	683,35	646,48	657,24	802,02	708,44	746,19	0,95	0,88
	Até 5º Ano Incompleto	951,82	794,95	865,58	956,27	856,88	906,08	0,84	0,90
	5º Ano Completo	937,84	813,47	873,08	1025,37	937,11	986,39	0,87	0,91
	6º ao 9º Ano Incompleto	878,60	780,81	823,45	1007,09	909,08	965,11	0,89	0,90
	Fundamental Completo	1040,19	887,18	967,62	1037,22	936,11	999,38	0,85	0,90
	Médio Incompleto	906,00	784,22	843,52	971,33	861,25	928,23	0,87	0,89
	Médio Completo	1141,78	941,18	1049,27	1232,94	1052,34	1162,89	0,82	0,85
	Superior Incompleto	1491,62	1224,22	1400,27	1920,73	1557,67	1823,58	0,82	0,81
	Superior Completo	3310,03	2281,92	3068,50	3989,13	2969,71	3784,74	0,69	0,74

Fonte: Elaboração própria. Microdados Censo Demográfico 2010, IBGE; Relação Anual de Informações Sociais 2010, MTE.

Nota: * Com expansão da Amostra

CAPÍTULO IV – APLICAÇÃO DE MODELOS DE EQUAÇÕES DE SALÁRIOS

A fim de executar uma comparação mais precisa entre as fontes de dados estudadas, neste capítulo serão produzidos e comparados modelos para a formação de renda dos trabalhadores. A primeira seção apresentará um modelo de regressão linear a ser utilizado individualmente nas duas bases; em seguida comparar-se-á um modelo para uma amostra mista delas.

IV.1 – EQUAÇÕES DE SALÁRIOS NO CENSO E RAIS.

Interessa-nos testar a relação funcional entre a característica de Cor ou Raça (d_i , 0 = pretos e pardos) e o logaritmo natural dos rendimentos do trabalho ($\ln y_i$), nos termos aqui expostos:

$$(1) \ln y_i = f(d_i), \quad i = 1, 2, \dots, n$$

Para tanto será utilizado um modelo de regressão linear múltipla por Mínimos Quadrados Ordinários (MQO). Sua forma será a de uma equação salário minceriana clássica:

$$(2) y = X\beta + \varepsilon$$

Em (2), β e y são vetores coluna de coeficientes angulares e rendimentos (em log natural), respectivamente; já X é a matriz de dados de individuais (inclusive d_i) e ε é o vetor de resíduo. Uma das condições necessárias para a execução da MQO é o da inexistência da correlação entre os resíduos e as variáveis independentes, isto é, que $E[\varepsilon|X] = 0$ ²⁰, assim a matriz de características X conterá uma série de variáveis de controle. São elas:

- (i) Cinco *dummies* para as diferentes regiões do Brasil – Sudeste, Sul, Norte, Centro-Oeste (sem Distrito Federal) e Distrito Federal. O Nordeste é tido como base, isto é, se as demais variáveis todas tiverem valor zero.

²⁰ Menezes-Filho (2002). Sabe-se, contudo, que para as fontes de dados estudadas a eliminação completa de covariância entre o resíduo e as variáveis dependentes é improvável, uma vez que há uma seleção escassa de variáveis de controle a serem utilizadas, ainda mais quando se compatibiliza questionários diferentes. Nesses casos, nem sempre é possível modificar a função utilizada ou introduzir novos controles. Por vezes, não o fazer é mesmo o mais desejável.

- (ii) *Dummy* para o sexo do indivíduo, valor zero para o sexo feminino.
- (iii) Oito *dummies* para nível educacional. Os grupos são aqueles mesmos discutidos na seção III.2. A base zero é formada por analfabetos.
- (iv) O log natural de idade, utilizado como *proxy* para a experiência profissional do indivíduo²¹.
- (v) *Dummy* para casos de cor ou raça não declarados na RAIS.

Os resultados da regressão estão contidos na Tabela 5 na página abaixo. É desnecessário salientar que, para todas as variáveis, os resultados são absolutamente significativos. Isso é consequência da amostra de aproximadamente 35 milhões de casos nas duas bases de dados (com expansão da amostra no Censo). A adequação da função à variável dependente, medida por R^2 , é de 0,414 para o Censo e 0,377 para a RAIS. Modestos, mas em linha com os resultados habituais para este tipo de modelo²². O resultado mais importante é o do coeficiente do *dummy* de Cor ou Raça: enquanto que no Censo há indicação de uma diferença da renda mediana atribuível à variável de cerca de 11,4%, na RAIS esta é de somente 6,4%. Das variáveis estudadas, a referente à Cor ou Raça representa o pior resultado em termos de diferença relativa (ver coluna “(a/b)”).

Igualmente, para as variáveis educacionais há grandes diferenças, maiores nos grupos de menor nível educacional, mas com semelhança nos retornos da população com nível superior incompleto e completo. Apesar da substancial diferença na desigualdade de renda por sexo, há similaridade em como são capturados pelos coeficientes nas duas fontes – é de cerca de 30,7% no Censo e 31,9% na RAIS. A elasticidade renda-idade medida pela função log-log na regressão com o logaritmo natural de idade é similar, 0,54 no Censo, 0,50 na RAIS. Há diferenças nas regiões geográficas do Sudeste e Distrito Federal. Na primeira, a renda é superestimada em relação ao Censo em 5%; no Distrito Federal é subestimada em 11,8%. A variável de característica racial não declarada apresenta um resultado interessante: é muito próxima de zero (embora estatística-

²¹ O mais comum para trabalhos sobre o tema é o uso de duas variáveis, idade e idade ao quadrado. Essa forma tem como base a experiência do mercado de trabalho norte-americano, a qual, a partir de aproximadamente 50 anos de idade, passa a revelar um declínio marginal da renda. Assim, a parábola formada descreve adequadamente esse comportamento. Entretanto, no mercado de trabalho brasileiro a relação é, no limite do aceitável, linear; os resultados (inclusive dessas amostras para a especificação quadrática) indicam declínio da renda somente aos 100 anos de idade – um resultado espúrio. Por fim, optou-se usar o logaritmo natural como foi feito em De Negri et al. (2001). Sobre o tema ver Lam e Levinson (1990).

²² Por exemplo, conferir Mincer (1974). Cain (1986) também mobiliza impressionante seleção de trabalhos sobre o tema.

mente significativo), o que sugere que a composição deste grupo seja semelhante ao da média da população²³.

Tabela 5

Resultados de um modelo de equações de salários minceriana por MQO. Comparação entre Censo e RAIS, Brasil, 2010.

CENSO 2010			RAIS 2010		Diferença CENSO/RAIS	
Variáveis	Coefficiente (a)	Valor T	Coefficiente (b)	Valor T	(a-b)	(a/b)
Intercepto	3,943^{oo}	3375,73	4,263^{oo}	2610,72	-0,320	0,925
β LN (LN-Idade)	,546 ^{oo}	2020,21	,501 ^{oo}	1717,65	0,045	1,090
β S (Sexo)	,307 ^{oo}	1746,41	,319 ^{oo}	1696,53	-0,012	0,963
β C (Cor)	,114 ^{oo}	635,53	,064 ^{oo}	307,98	0,050	1,785
β NO (Norte)	,141 ^{oo}	324,02	,145 ^{oo}	301,06	-0,004	0,970
β SE (Sudeste)	,225 ^{oo}	944,85	,275 ^{oo}	1035,15	-0,050	0,819
β SU (Sul)	,188 ^{oo}	635,94	,204 ^{oo}	621,02	-0,016	0,920
β DF (D. Federal)	,363 ^{oo}	533,76	,245 ^{oo}	357,52	0,118	1,479
β CO (Centro-Oeste)	,179 ^{oo}	451,63	,139 ^{oo}	313,22	0,040	1,291
β 1 (Primário Incompleto)	,207 ^{oo}	333,10	,140 ^{oo}	107,73	0,066	1,474
β 2 (Primário Completo)	,229 ^{oo}	357,12	,180 ^{oo}	139,86	0,050	1,276
β 3 (Ginásio Incompleto)	,304 ^{oo}	495,28	,222 ^{oo}	176,91	0,083	1,372
β 4 (Fundamental Completo)	,406 ^{oo}	664,84	,257 ^{oo}	207,16	0,149	1,581
β 5 (Médio Incompleto)	,412 ^{oo}	659,08	,288 ^{oo}	229,38	0,124	1,431
β 6 (Médio Completo)	,547 ^{oo}	949,87	,464 ^{oo}	378,47	0,082	1,178
β 7 (Superior Incompleto)	,814 ^{oo}	1288,46	,853 ^{oo}	657,76	-0,039	0,954
β 8 (Superior Completo)	1,347 ^{oo}	2246,27	1,329 ^{oo}	1064,32	0,018	1,014
β lg (Ignorados RAIS)	-	-	,008 ^{oo}	19,19	-	-
Nº Observações	37.089.379		34.924.625			
	R	R²	R²-Aj.	R	R²	R²-Aj.
	0,643	0,414	0,414	0,614	0,377	0,377
Análise de Variância						
Modelo	CENSO			RAIS		
	Regressão	Resíduos	Total	Regressão	Resíduos	Total
Soma dos Quadrados	6.745.153	9.564.334	16.309.487	5.859.638	9.672.503	15.532.142
Graus de Liberdade	16	37.089.362	37.089.378	17	34.924.608	34.924.625
Média dos Quadrados	421.572	0,258		344.685	0,277	
F	1.634.807			1.244.556		
Significância de F	< 0,0001			< 0,0001		

** Significante a 1%

Fonte: Elaboração Própria. Microdados Censo Demográfico 2010, IBGE; Relação Anual de Informações Sociais 2010, MTE. Com expansão da amostra.

²³ De fato, no processo iterativo de construção dos modelos, quanto mais variáveis geográficas incorporadas, menor o valor do coeficiente β lg, mostrando indiretamente a presença de multicolinearidade (quicã relação causal entre os dois).

IV.2 – MODELO DE EQUAÇÕES DE SALÁRIOS EM UMA AMOSTRA MISTA.

Para testar de modo mais rigoroso as diferenças entre as bases de dados, o modelo aqui descrito vai testar, simultaneamente, os retornos sobre as características testadas na seção anterior e o valor das diferenças entre as duas. Desta forma é possível verificar, através do teste t, a adequação destas variáveis entre nossas fontes de dados; a hipótese nula é de que não há diferenças nos parâmetros. Para tanto, foi selecionada uma amostra aleatória pequena do Censo e da RAIS, de 32.000 casos cada. Eles são pertencentes à mesma população estudada antes, mas estão ausentes os casos de cor não informada na RAIS. A partir desta nova base, o seguinte modelo será estimado:

$$(3) \ y = X\beta_C + (d_{Ri}X)\beta_D + \varepsilon$$

d_{Ri} é uma variável binária escalar para marcar o pertencimento do i -ésimo caso da amostra a uma ou outra base de dados, com o valor positivo representando casos retirados da RAIS. $d_{Ri}X$ forma novas variáveis de características específicas à RAIS. Obtém-se da equação (3), dado $E[y|d_{Ri} = 0, X]$:

$$(4) \ y = X\beta_C + \varepsilon$$

Logo, (4) se trata da equação de salários minceriana (2) somente para o Censo Demográfico. β_C é, portanto, o vetor coluna de coeficientes angulares da equação salários desta da base. Agora, se $E[y|d_{Ri} = 1, X]$, tem-se que:

$$(5) \ y = X(\beta_C + \beta_D) + \varepsilon$$

Logo, (5) é a equação minceriana para a RAIS. $\beta_C + \beta_D$ é igual ao vetor de coeficientes angulares de características X para a amostra da RAIS. Chamemo-la de β_R . Segue-se que β_D representa a diferença nos coeficientes de retorno financeiro das características estudadas para o Censo e a RAIS que serão testados, ou seja, um vetor de coeficientes angulares diferenciais. Interessa-nos fazer o teste de hipóteses sobre os valores de β_D , confirmando ou rejeitando as diferenças entre as bases de dados. Os resultados estão contidos na tabela 6; os parâmetros com subscrito “R” designam os coeficientes das variáveis $d_{Ri}X$.

Tabela 6

Resultados de um modelo de equações de salários por MQO em uma amostra mista. Comparação entre Censo e RAIS, Brasil, 2010.

Variáveis	Intercepto	β_{LN}	β_S	β_C	β_{NO}	β_{SE}	β_{SU}	β_{DF}	β_{CO}	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Coefficiente	4,071	0,503	0,318	0,099	0,182	0,228	0,199	0,389	0,221	0,182	0,234	0,308	0,412	0,423	0,537	0,770	1,261
Valor T	105,04	54,45	52,61	16,12	12,01	27,33	20,47	12,02	17,21	9,77	12,15	16,51	22,11	22,06	30,99	39,18	68,51
P-Valor	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

Variáveis	β_{BR} (Base)	β_{RLN}	β_{RS}	β_{RC}	β_{RNO}	β_{RSE}	β_{RSU}	β_{RDF}	β_{RCO}	β_{R1}	β_{R2}	β_{R3}	β_{R4}	β_{R5}	β_{R6}	β_{R7}	β_{R8}
Coefficiente	0,200 ^{°°}	-0,013	0,012	-0,023 ^{°°}	-0,059 ^{°°}	0,038 ^{°°}	-0,010	-0,171 ^{°°}	-0,077 ^{°°}	-0,032	-0,052	-0,050	-0,124 ^{°°}	-0,111 ^{°°}	-0,052	0,120 ^{°°}	0,110 ^{°°}
Valor T	3,11	-1,02	1,43	-2,54	-2,72	3,18	-0,67	-4,40	-3,99	-0,72	-1,17	-1,15	-2,89	-2,53	-1,23	2,66	2,54
P-Valor	0,2%	31,0%	15,3%	1,1%	0,7%	0,1%	50,5%	0,0%	0,0%	46,9%	24,2%	24,9%	0,4%	1,1%	21,8%	0,8%	1,1%

Nº Observações		64.000	Análise de Variância	Soma dos Quadrados	Graus de L.	Média dos Quadrados	F	Sig. de F	
R		0,626		Regressão	10.586	33	320,78	1.252	< 0,001
R²		0,392		Resíduos	16.386	63.966	0,256		
R²-Aj.		0,392		Total	26.972	63.999			

[°] Significativo para p-valor inferior a 5%

^{°°} Significativo para p-valor inferior a 1%

Fonte: Elaboração Própria. Microdados Censo Demográfico 2010, IBGE; Relação Anual de Informações Sociais 2010, MTE. Sem expansão da amostra.

Dos resultados da tabela 6, confirma-se a hipótese nula somente para sete das 16 variáveis diferenciais. O valor do coeficiente angular diferencial (β_D) da variável sexo não é estatisticamente significativo a 5% (p-valor: 15,3%); já a variável de cor apresenta diferença pronunciada – de 2,3 pontos percentuais na renda do trabalho mediana –, sendo significativa para o p-valor do teste t nesta tolerância. Nas variáveis geográficas do país também há considerável diferença no valor da renda mediana atribuível às mesmas. Em três regiões e no Distrito Federal, o coeficiente angular diferencial é significativo para um nível de 1%, com exceção da região Sul em que é insignificante. Nas variáveis educacionais, metade dos grupos apresentaram resultados significativos para 5%. São eles: a população com ensino fundamental completo, médio incompleto, superior incompleto e completo.

Portanto, os resultados obtidos não confirmam nossas expectativas. Não somente existem diferenças significativas em boa parte das variáveis, como é particularmente grande no caso de Cor ou Raça. O sinal do coeficiente é negativo. Isto significa que, em relação ao Censo Demográfico, a RAIS subestima o efeito da discriminação racial na determinação dos rendimentos no mercado de trabalho formal brasileiro. Isso pode determinar para muitas agendas de pesquisa a exclusão desta fonte de dados, uma vez que sustenta com dificuldade a hipótese de discriminação racial. Há, ainda por cima, dificuldades adicionais em relação às especificações dos modelos utilizados: uma vez que há, inevitavelmente, uma correlação não desprezível entre as variáveis de controle do modelo e a Cor ou Raça do indivíduo, a heterogeneidade na diferença da composição comparada por região entre as duas fontes de dados produz resultados que, no caso da RAIS, são pouquíssimo confiáveis. Ou seja, não se observa plenamente o fenômeno de embranquecimento, e embora aponte a presença de discriminação racial, não o é em grau maior. A negação do primeiro fato estilizado parece comprometer os resultados das análises paramétricas²⁴.

²⁴ Uma evidência disso pode ser encontrada em um modelo alternativo (não exibido nesta monografia). Na ausência de variáveis de controle geográfico, não há alteração de nenhum dos resultados dos testes de hipótese e pequeníssimas dos coeficientes angulares diferenciais encontrados, **exceto** para a variável de cor ou raça. É tão flagrante que, sem este controle, o valor de β_{RC} é de somente -0,1% da renda mediana, com o incrível p-valor de 92,3%. Expressa-se aqui evidência de multicolinearidade. Isto também é verdadeiro para o conteúdo do capítulo V. **Disto resulta a conclusão importante de que a base pode ser imprestável para estudos nacionais, mas não para regiões geograficamente mais restritas como municípios e UFs.**

CAPÍTULO V – DECOMPOSIÇÃO DE OAXACA PARA AS EQUAÇÕES DE SALÁRIOS

Neste capítulo final serão exibidos os resultados da decomposição das equações de salários conforme o modelo do capítulo anterior. Na primeira seção explica-se o processo, na segunda são discutidos os resultados por meio de tabelas.

V.1 – DECOMPOSIÇÃO DE OAXACA.

Conforme o simples método desenvolvido por Oaxaca (1973), os modelos do capítulo anterior podem ser decompostos em dois termos, de modo a identificar a proporção do diferencial de renda explicada pela presença de discriminação. Voltando à equação (2), suponha-se que a formação de salários médios da população masculina branca, que definimos como o da experiência padrão, seja:

$$(6) \ln \bar{y}_{MB} = \bar{X}_{MB}' \beta_{MB}$$

β_{MB} e \bar{X}_{MB} são os respectivos vetores coluna de coeficientes angulares e de variáveis de características médias desta população. Portanto, os três grupos restantes, - a saber: pretos e pardos do sexo masculino (*MP*), brancas e pretas e pardas do sexo feminino (*FB*, *FP*, respectivamente) - na ausência de discriminação (definido de modo restrito ao modelo) tem como equação salário:

$$(7) \ln \bar{y} = \bar{X}_i' \beta_{MB} \quad i = MP, FB, FP$$

O diferencial de renda entre um desses grupos e o padrão é dado por:

$$(8) \ln \bar{y}_{MB} / \bar{y}_i = \bar{X}_{MB}' \beta_{MB} - \bar{X}_i' \beta_i$$

Que pode ser transformado da seguinte forma:

$$(9) \bar{X}_{MB}' \beta_{MB} - \bar{X}_i' \beta_i + \bar{X}_i' \beta_{MB} - \bar{X}_i' \beta_{MB}$$

$$(10) (\bar{X}_{MB} - \bar{X}_i)' \beta_{MB} + \bar{X}_i' (\beta_{MB} - \beta_i)$$

O termo à esquerda descreve o diferencial de renda associado às diferenças nos vetores de características que são, na execução dos modelos dessa monografia, as dife-

renças na idade (como *proxy* de experiência), região geográfica e educação. O termo à direita descreve as diferenças causadas pelo diferencial de remuneração dessas características, associadas aos coeficientes angulares dos vetores β_{MB} e β_i . Cada um desses termos compõem um total, do qual podemos sinteticamente retirar uma porcentagem demonstrativa do tamanho e importância da discriminação racial e de gênero no mercado de trabalho em relação, no caso, à população masculina e branca.

V.2 – RESULTADOS DA DECOMPOSIÇÃO.

Para estimar as equações (7) e (8), foram novamente produzidas regressões lineares múltiplas por MQO. Os resultados daí obtidos estão reproduzidos nas tabelas 7 e 8, respectivamente, para o Censo Demográfico e para a RAIS em 2010. Nelas se encontram resultados equivalentes aos da seção IV.1. Com a possibilidade de observar os parâmetros encontrados nessas regressões, verifica-se a diferença de remuneração entre os grupos de sexo e raça por meio de comparação direta, e não através das variáveis instrumentais utilizadas.

Na RAIS, assim como de modo geral para os grupos que não os de brancos do sexo masculino, o intercepto possui valor maior, refletindo não só diferenças na remuneração do grupo base, como também a maior inadequação das demais variáveis de controle na determinação da renda – espelhado por um R^2 menor nesses grupos. Assim, nota-se que há diferenças grandes na remuneração por idade e por educação entre os grupos estudados; mas a diferença é ainda maior entre as duas pesquisas. Por exemplo: enquanto que com ensino médio completo a mediana de remuneração é aproximadamente 51,6% maior para mulheres brancas e 67,6% para homens brancos, estes valores são de apenas 39,2% e 51,5%, respectivamente, na RAIS.

O resultado da decomposição destes números pode ser observado na tabela 9, três páginas adiante. Nela está contido o log-valor das diferenças do termo de produtividade ou explicado pelo modelo $((\bar{X}_{MB} - \bar{X}_i)' \beta_{MB})$ e no outro o termo de discriminação ou residual $(\bar{X}_i' (\beta_{MB} - \beta_i))$, assim como a porcentagem explicada pelo último. A discriminação para o modelo especificado é maior para a população branca e de sexo feminino. O valor transpassa o cem por cento, uma vez que o termo de produtividade é negativo. Isto significa dizer que, na ausência de discriminação, este grupo deveria receber

medianamente mais, uma vez que o mesmo possui uma composição de população com nível educacional mais avançado. Este valor é de 167,9% no Censo e 139,2% na RAIS, indicando (neste caso) uma composição educacional mais semelhante na RAIS, mas ainda assim um nível de discriminação quase idêntico.

Tabela 7
Resultados de um modelo de equações de rendimentos por MQO para quatro populações. Censo 2010, Brasil.

População	Masculina Branca	Masculina Preta e Parda	Feminina Branca	Feminina Preta e Parda
Variáveis				
Intercepto	3,699^{oo} (1588,10)	4,402^{oo} (2411,18)	4,313^{oo} (1241,49)	4,829^{oo} (1784,02)
β_{LN} (LN-Idade)	,692 ^{oo} (1336,79)	,510 ^{oo} (1118,28)	,486 ^{oo} (812,19)	,353 ^{oo} (593,03)
β_{NO} (Norte)	,157 ^{oo} (136,42)	,139 ^{oo} (239,26)	,136 ^{oo} (101,12)	,128 ^{oo} (168,42)
β_{SE} (Sudeste)	,258 ^{oo} (458,20)	,216 ^{oo} (627,64)	,215 ^{oo} (338,76)	,157 ^{oo} (355,26)
β_{SU} (Sul)	,230 ^{oo} (376,85)	,218 ^{oo} (387,88)	,143 ^{oo} (209,63)	,141 ^{oo} (201,36)
β_{DF} (D. Federal)	,416 ^{oo} (258,30)	,288 ^{oo} (275,52)	,456 ^{oo} (280,22)	,315 ^{oo} (263,12)
β_{CO} (Centro-Oeste)	,210 ^{oo} (233,64)	,218 ^{oo} (383,26)	,095 ^{oo} (89,33)	,115 ^{oo} (150,06)
β_1 (Primário Incompleto)	,261 ^{oo} (192,58)	,179 ^{oo} (233,88)	,246 ^{oo} (91,72)	,152 ^{oo} (96,83)
β_2 (Primário Completo)	,293 ^{oo} (211,81)	,221 ^{oo} (274,76)	,208 ^{oo} (76,75)	,154 ^{oo} (95,19)
β_3 (Ginásio Incompleto)	,404 ^{oo} (299,19)	,290 ^{oo} (382,55)	,259 ^{oo} (97,06)	,190 ^{oo} (123,06)
β_4 (Fundamental Completo)	,510 ^{oo} (382,41)	,375 ^{oo} (483,57)	,389 ^{oo} (148,23)	,291 ^{oo} (190,38)
β_5 (Médio Incompleto)	,526 ^{oo} (382,41)	,362 ^{oo} (453,87)	,414 ^{oo} (156,47)	,284 ^{oo} (183,58)
β_6 (Médio Completo)	,676 ^{oo} (528,80)	,505 ^{oo} (712,31)	,516 ^{oo} (201,01)	,388 ^{oo} (267,73)
β_7 (Superior Incompleto)	,971 ^{oo} (710,63)	,786 ^{oo} (857,29)	,767 ^{oo} (294,04)	,601 ^{oo} (389,93)
β_8 (Superior Completo)	1,576 ^{oo} (1208,13)	1,272 ^{oo} (1415,37)	1,280 ^{oo} (498,38)	1,027 ^{oo} (688,01)
Nº Observações	11.514.014	11.002.957	8.358.041	5.756.043
R	0,659	0,550	0,627	0,558
R²	0,434	0,302	0,393	0,312
Significância de F	< 0,001	< 0,001	< 0,001	< 0,001

^{oo} Significante a 1%

Fonte: Elaboração Própria. Microdados Censo Demográfico 2010, IBGE. Com expansão da amostra.

Tabela 8
Resultados de um modelo de equações de rendimentos por MQO para quatro populações. RAIS, 2010, Brasil.

População	Masculina Branca	Masculina Preta e Parda	Feminina Branca	Feminina Preta e Parda
Variáveis				
Intercepto	4,165^{°°} (1336,12)	4,446^{°°} (1665,27)	4,882^{°°} (995,82)	5,089^{°°} (1072,44)
β_{LN} (LN-Idade)	,613 ^{°°} (1185,51)	,536 ^{°°} (932,46)	,370 ^{°°} (618,27)	,319 ^{°°} (439,91)
β_{NO} (Norte)	,163 ^{°°} (120,95)	,160 ^{°°} (241,84)	,130 ^{°°} (83,60)	,116 ^{°°} (140,06)
β_{SE} (Sudeste)	,320 ^{°°} (514,06)	,278 ^{°°} (678,74)	,252 ^{°°} (355,64)	,178 ^{°°} (358,54)
β_{SU} (Sul)	,245 ^{°°} (365,45)	,251 ^{°°} (280,93)	,173 ^{°°} (228,76)	,161 ^{°°} (154,96)
β_{DF} (D. Federal)	,232 ^{°°} (168,93)	,217 ^{°°} (177,00)	,312 ^{°°} (205,50)	,230 ^{°°} (166,28)
β_{CO} (Centro-Oeste)	,170 ^{°°} (174,71)	,182 ^{°°} (256,42)	,089 ^{°°} (78,78)	,093 ^{°°} (104,55)
β_1 (Primário Incompleto)	,142 ^{°°} (55,40)	,165 ^{°°} (94,20)	,058 ^{°°} (12,76)	,054 ^{°°} (13,01)
β_2 (Primário Completo)	,200 ^{°°} (79,65)	,215 ^{°°} (121,68)	,086 ^{°°} (19,37)	,071 ^{°°} (17,49)
β_3 (Ginásio Incompleto)	,261 ^{°°} (105,39)	,250 ^{°°} (146,20)	,120 ^{°°} (27,42)	,093 ^{°°} (23,55)
β_4 (Fundamental Completo)	,298 ^{°°} (121,57)	,272 ^{°°} (160,93)	,174 ^{°°} (39,98)	,137 ^{°°} (34,89)
β_5 (Médio Incompleto)	,330 ^{°°} (133,02)	,283 ^{°°} (163,64)	,232 ^{°°} (53,28)	,166 ^{°°} (42,07)
β_6 (Médio Completo)	,515 ^{°°} (211,40)	,453 ^{°°} (273,23)	,392 ^{°°} (90,71)	,330 ^{°°} (84,99)
β_7 (Superior Incompleto)	,936 ^{°°} (366,52)	,811 ^{°°} (404,45)	,765 ^{°°} (174,85)	,654 ^{°°} (163,45)
β_8 (Superior Completo)	1,454 ^{°°} (589,26)	1,271 ^{°°} (695,65)	1,220 ^{°°} (281,75)	1,088 ^{°°} (277,28)
Nº Observações	12.501.686	7.962.772	8.358.041	3.997.222
R	0,605	0,529	0,617	0,575
R²	0,366	0,280	0,381	0,330
Significância de F	< 0,001	< 0,001	< 0,001	< 0,001

°° Significante a 1%

Fonte: Elaboração Própria. Microdados Relação Anual de Informações Sociais 2010, MTE.

Para a população preta e parda masculina, não há similaridade nos dois termos e na porcentagem explicada. Somente 31,6% das diferenças salariais entre essas populações é explicada pela presença de discriminação no Censo; na RAIS o valor é ainda

menor, sendo estimado em 20,1%. Logo, sem similaridades. Contudo, para a população de sexo feminino preta e parda, os números são semelhantes: 86,3% na primeira fonte e 80,3% na segunda. Isso reflete, possivelmente, a predominância da discriminação de gênero sobre a racial no modelo adotado. Novamente, o resultado parece ser de menor detecção de discriminação na RAIS.

Tabela 9

Resultados da decomposição de Oaxaca para três grupos em relação à população masculina branca, em log natural e (%). Comparação entre Censo e RAIS, Brasil, 2010.

População	Diferença - Termo de Produtividade (a)	Diferença - Termo de Discriminação (b)	Porcentagem explicada por discriminação [b/(a/b)]
CENSO			
Masculina Preta e Parda	0,214	0,099	31,6%
Feminina Branca	-0,136	0,337	167,9%
Feminina Preta e Parda	0,065	0,408	86,3%
RAIS			
Masculina Preta e Parda	0,192	0,048	20,1%
Feminina Branca	-0,093	0,330	139,2%
Feminina Preta e Parda	0,090	0,364	80,3%

Fonte: Elaboração própria

CONCLUSÃO

Vimos a partir do capítulo segundo que não é tarefa banal compatibilizar as duas fontes de dados. De fato, isto não é plenamente possível. No que pese as diferenças sazonais, geográficas e metodológicas, a maior de todas elas é, com certeza, a do informante. Ora, os modelos que utilizamos a partir do terceiro capítulo são embasados nos marcos teóricos clássicos que procuram executar a tarefa hercúlea de estimar a formação dos rendimentos do trabalho predados a características abstratas e, certamente, apenas indiretamente mensuráveis. Esse é o caso com as variáveis educacionais e de idade, sobre as quais presumimos representarem a produtividade do trabalhador. Agora, o que fazer do fenótipo ou raça do indivíduo? Que valor atribuir a medidas diferentes sobre um mesmo fenômeno? Se forem diferentes, não podem estar ambas corretas? Bem, neste caso até poderiam.

Esperávamos, nessa comparação, observar a existência de dois fatos: o embranquecimento da população heteroatribuída e a presença de maior discriminação racial na mesma. O embranquecimento foi constatado, mas não é fato conclusivo e nem regular no Brasil. A RAIS não é comparável com as pesquisas da PESB, PCERP ou Datafolha, mas apesar disso, pode apontar problemas com pesquisas utilizando estas fontes se as divergências apontadas forem de cunho metodológico. Elas trabalham com amostras pequenas de pesquisas nacionais, e suas análises se baseiam sobre estes dados na sua totalidade, sem controle minucioso das variáveis geográficas. Telles e Lim (1998), por exemplo, introduzem apenas dois *dummies* no teste econométrico efetuado: para o Nordeste e áreas urbanas, com o único propósito de separá-las pela sua formação salarial diferenciada. A dificuldade, como vimos, é que o processo em que há embranquecimento e, portanto, discordância nos critérios de classificação, é tão fortemente determinado por área geográfica que podemos chegar a observações falsas dos coeficientes de discriminação racial. Logo, chega-se a conclusões equivocadas. É claro que se pronuncia tais desfechos sem sentenças - seria necessário investigar essas outras bases de dados para determinar se coincidem com a RAIS (a falsidade pode ser facilmente atribuída a ela).

Os resultados dos modelos dos capítulos IV e V rejeitam a existência de discriminação racial maior. Pelo contrário, mostram uma diferença considerável: β_C , o coefi-

ciente angular da variável *dummy* racial, indica que brancos recebem 11,4% a mais no Censo, mas somente 6,4% na RAIS. Já o coeficiente β_{RC} , na seção IV.2, indica uma diferença estatisticamente significativa de 2,3%. Essa última diferença só é passível de ser observada quando são controladas as regiões geográficas. Portanto, para muitas agendas de pesquisa, as estimativas inferiores já são suficientes para rejeitarem o uso completo da RAIS para o estudo das desigualdades raciais. Vimos ainda, no capítulo V, que a variável tem menor poder explicativo nos diferenciais de renda observados na RAIS. Apesar disto, as duas bases não permitem descartar, a partir dos modelos e controles utilizados, a presença de discriminação racial no mercado de trabalho formal brasileiro.

Este último resultado é, não obstante, encorajador. A partir dele podemos concluir que é adequada para os fins expressos na introdução deste texto, a saber: para estudos empíricos sobre discriminação racial e avaliação de políticas públicas. Segue, portanto, que a linha de pesquisa aqui iniciada ainda está em aberto. Os modelos utilizados na qual se baseiam essas conclusões, são, indubitavelmente, simples e possuem deficiências óbvias para além daquelas mencionadas. Por exemplo, o método por MQO – em conjunto com as variáveis binárias utilizadas – só nos dá o resultado de parâmetros para a população média, desconsiderando que a variância nos salários é proporcionalmente maior ao longo dos quantis de renda, logo com possível discriminação mais ampla ao longo deles (SOARES, 2000; BAILEY, LOVEMAN & MUNIZ, 2013). Igualmente o modelo apresentado na seção IV.2, nos permite observar a interação das variáveis empregadas e as fontes de dados provenientes mas não entre as variáveis em si. Isto é, não observamos como varia o coeficiente angular ao qual atribuímos o peso da discriminação racial em diferentes regiões geográficas, níveis educacionais e de idade. E finalmente há ainda uma série de variáveis a serem testadas como os de grupo ocupacional da CBO e de setores de atividade, que possuem relevância considerável na determinação da renda individual assim como da de populações (dado composições diferenciadas).

Da parte deste autor, a conclusão final a ser tirada é a seguinte: as duas fontes de dados comparadas não são iguais ou sequer parecidas. O abismo metodológico é largo. Contudo, lembremos não ser este o propósito da presente monografia. Procuramos validar a utilização da variável racial na RAIS com base na autoridade de outra pesquisa, neste caso, o Censo Demográfico de 2010. Faz-se as ressalvas: não se indica a base para

o uso de estudos sobre grandes agregados nacionais; para isto basta usar a PNAD, por exemplo, e sem prejuízo por deficiência de amostra. Para estudos em setores reduzidos e localizados, pode muito bem ainda ser adequado e talvez seja a fonte unicamente disponível para o tema. De resto, recordemos que, embora no caso dos registros administrativos, o pesquisador não seja agraciado por questionários e amostra de *designs* superiores, tais registros estão sempre sujeitos a melhorias e inovações técnicas. Portanto, fiquemos com um olho atento, em especial para uma futura (mas improvável) introdução de dados sobre o setor público. Porém, desde já a RAIS pode ser importante para a avaliação de políticas públicas de combate à discriminação racial.

REFERÊNCIAS BIBLIOGRÁFICAS

ARIAS, A. R.; CORDEIRO, S. H. T. C. Uma discussão sobre a produção e uso de dados sobre o mercado de trabalho. *Revista Brasileira de Estudos da População*, Campinas, v. 7, p. 219-235, 1990.

ARROW, K. Higher education as a filter. *Journal of Public Economics*, Harvard University, U.S.A, v. 2, n. 3, 1973.

AZEVEDO, B. *O emprego no Rio Grande do Sul: RAIS, PNAD e PME*. Porto Alegre: FEE, 1985.

BAILEY, S. R.; LOVEMAN, M.; MUNIZ, J. O. Measures of Race and the analysis of racial inequality in Brazil. *Social Science Research* (Print), [s.l.] , v. 42, p. 106-119, 2013.

BARROS, R. P.; ULYSSEA, G.; CURY, S. A desigualdade de renda encontra-se subestimada? Uma análise comparativa usando a PNAD, POF e Contas Nacionais. In: BARROS, R. P.; FOGUEL, M.; ULYSSEA, G. (Org.). *Desigualdade de Renda no Brasil: uma análise da queda recente*. [s.l.] IPEA, 2007.

CAIN, G. G. The economic analysis of labor market discrimination: a survey. In: ASHENFELTER, O.; LAYARD, R. (eds.). *Handbook of Labor Economics*. North-Holand, 1986.

CORSEUIL, C. H. L.; COELHO, A. M. Diferenciais Salariais no Brasil: um breve panorama. In: CORSEUIL, C. H. L.; SANTOS, D. D.; FERNANDES, R.; MENEZES-FILHO, N.; COELHO, A. M. *Estrutura Salarial: Aspectos Conceituais e novos resultados para o Brasil*. Rio de Janeiro: IPEA, 2002.

CORSEUIL, C. H. L.; SANTOS, D. D. Fatores que determinam o nível salarial no setor formal brasileiro. In: CORSEUIL, C. H. L.; SANTOS, D. D.; FERNANDES, R.; MENEZES-FILHO, N.; COELHO, A. M. *Estrutura Salarial: Aspectos Conceituais e novos resultados para o Brasil*. Rio de Janeiro: IPEA, 2002.

DEGLER, C. N. *Neither Black nor White: Slavery and Race Relations in Brazil and the United States*. Madison: University of Wisconsin Press, 1971.

DE NEGRI, J. A.; CASTRO, P. F.; SOUZA, N. R.; ARBACHE, J. S. *Mercado formal de trabalho: comparação entre os microdados da RAIS e da PNAD* (Texto para Discussão nº 840). Brasília: IPEA, 2001.

GOLDBERGER, A. S. *A Course in Econometrics*. Cambridge: Harvard University Press, 1991.

HASENBALG, C. *Discriminação e desigualdades raciais no Brasil*. Rio de Janeiro: Graal, 1979.

IPARDES. *Avaliação do alcance e limites da RAIS e a Lei 4923/65*. Curitiba, 1987.

JANNUZZI, P. M. As Potencialidades Analíticas da RAIS para Estudos sobre a Estrutura do Mercado Formal de Trabalho. In: ENCONTRO NACIONAL DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS DO TRABALHO, 4, 1994.

_____. Fontes de informação sócio-demográfica para planejamento no setor público. *Revista de Administração Pública* (Impresso), Rio de Janeiro, v. 29, n. 3, p. 197-210, 1995.

_____. *Indicadores Sociais no Brasil: conceitos, fontes de dados e aplicações*. Campinas: Editora Alínea/PUC-Campinas, 2001.

KOENKER, R., BASSET, G. Regression quantiles. *Econometrica*, New York, v. 46, n. 33-50, 1978.

LAM, D.; LEVINSON, D. Idade, experiência e diferenciais de renda no Brasil. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 20, n. 2, 1990.

MARTINS, R. *Desigualdades e discriminação de gênero e raça no mercado de trabalho brasileiro no final do século XX*. Brasília: OIT, 2003a.

MARTINS, R. *Desigualdades raciais e políticas de inclusão racial: um sumário da experiência brasileira recente*. [s.l.]: CEPAL, 2003b.

MENEZES-FILHO, N. Equações de rendimentos: questões metodológicas. In: CORSEUIL, C. H. L.; SANTOS, D. D.; FERNANDES, R.; MENEZES-FILHO, N.; COELHO, A. M. *Estrutura Salarial: Aspectos Conceituais e novos resultados para o Brasil*. Rio de Janeiro: IPEA, 2002.

MINCER, J. *Schooling, experience and earnings*. New York: Columbia University Press, 1974.

MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO; INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. *Características Étnico-raciais da População: um estudo das categorias de classificação de cor ou raça 2008*. Rio de Janeiro, 2011.

_____. *Estudos e tratamento da variável rendimento no censo demográfico 2010*. Março, 2012. Acesso em: 9 de novembro de 2014. Disponível em: <ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_Gerais_da_Amostra/Estudo_e_tratamento_rendimentos.pdf>

_____. Metodologia do Censo Demográfico 2010. *Série Relatórios Metodológicos*, v. 41, jan., 2013.

MINISTÉRIO DO TRABALHO E EMPREGO (1999). Portaria nº1.740/99. Brasília, 1999. Acesso em: 3 de novembro de 2014. Disponível em: <http://portal.mte.gov.br/data/files/FF8080812BE914E6012BEECC06A8760A/p_19991026_1740.pdf>.

_____. *Manual de Orientação RAIS; ano-base 2010*. Brasília, 2011. Acesso em: 3 de novembro, 2014. Disponível em: <http://www.rais.gov.br/rais_ftp/ManualRAIS2010.pdf>.

_____. *Nota Técnica MTE 093/2014*. Brasília, 2014. Acesso em: 3 de novembro, 2014. Disponível em:

<http://portal.mte.gov.br/data/files/FF808081475961470147CC2F62B03900/Nota%20T%C3%A9cnica%20MTE%202013_2.pdf>.

NATIONAL RESEARCH COUNCIL. (2004). *Measuring Racial Discrimination. Panel on Methods for Assessing Discrimination*. Rebecca M. Blank, Marilyn Dabady, and Constance F. Citro, Editors. *Committee on National Statistics, Division of Behavioral and Social Sciences and Education*. Washington, DC: *The National Academies Press*.

OLIVEIRA, L.; PORCARO, R. M.; COSTA, T. (s/d). O lugar do negro na força de trabalho. Rio de Janeiro: Centro de Estudos Afro-Asiáticos / Conjunto Universitário Cândido Mendes.

OSÓRIO, R. G. O sistema classificatório de cor ou raça do IBGE. In: BERNARDINO, J.; GALDINO, D. (Org.). *Levando a raça a sério: ação afirmativa e a universidade*. Rio de Janeiro: DP&A, LPP, UERJ, 2004.

PAIXÃO, M.; SILVA, G. M. D. Mixed and Unequal: new perspectives on Brazilian ethnoracial relations. In: TELLES, E. (Org.). *Pigmentocracies: ethnicity, race, and color in Latin America*. Chapel Hill/North Carolina: The University of North Carolina Press, 2014.

PAIXÃO, M.; ROSSETTO, I. Levantamento das fontes de dados estatísticos sobre a variável cor ou raça no Brasil contemporâneo: terminologias classificatórias, qualidade das bases de dados e implicações para as políticas públicas. In: ENCONTRO ANUAL DA ANPOCS, 35, 2011.

PAIXÃO, M.; ROSSETTO, I.; MONCORES, E.; CÂMARA, G. C. P. Investigação sobre qualidade da variável cor ou raça na RAIS através de um estudo comparativo com a PNAD do IBGE. In: Encontro Anual da ANPOCS, 36, 2012.

PRADO, P. *Retrato do Brasil: Ensaio sobre a tristeza brasileira*, Novembro de 1928. Disponível em: <<http://www.ebooksbrasil.org/eLibris/pauloprado.html>>

RAMOS, L. O Desempenho Recente do Mercado de Trabalho. *Revista de Economia Política* (Impresso), São Paulo, v. 29, p. 406-420, 2009.

ROSEMBERG, F.; PIZA, E. Cor nos censos brasileiros. *Revista USP*, São Paulo, nº 40, p. 122-127, 1998.

SABÓIA, J. L. M., TOLIPAN, R. M. L. A Relação Anual de Informações Sociais (RAIS) e o Mercado Formal de Trabalho no Brasil. Rio de Janeiro: IEI/UFRJ, 1985.

SCHULTZ, T. W. *O Valor Econômico da Educação*. Rio de Janeiro: Zahar, 1973.

SOARES, S. *O perfil da discriminação no mercado de trabalho: homens negros, mulheres brancas e mulheres negras* (Texto para discussão n. 769). Brasília: IPEA, 2000

SOUZA, P. H. G. F. *A distribuição de renda nas pesquisas domiciliares brasileiras: harmonização e comparação entre Censos, PNADs e POFs* (Texto para discussão n. 1832). Brasília: IPEA, 2013.

SPENCE, M. Job market signalling. *Quarterly Journal of Economics*, Harvard University, U.S.A, v. 87, n. 3, 1973.

TELLES, E. *Racismo à brasileira: uma nova perspectiva sociológica*. Rio de Janeiro: Relume Dumará, 2003.

TELLES, E. E.; LIM, N. Does it matter who answers the race question? Racial classification and income inequality in Brazil. *Demography*, [s.l.], v. 35, n. 4, nov., 1998.

VALLE SILVA, N. Uma nota sobre “raça social” no Brasil. In: HASENBALG, C.; VALLE SILVA, N.; LIMA, M. (Orgs.). *Cor e estratificação social*. Rio de Janeiro: Contra Capa Livraria, 1999.

VALLE, N. O preço da cor: diferenciais raciais na distribuição da renda no Brasil. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 10, n. 1, abr., p. 21-44, 1980.